Nowcasting Real Economic Activity in the Euro Area: Assessing the Impact of Qualitative Surveys

Raïsa Basselier[*], David de Antonio Liedo[†], Geert Langenus[‡]

Abstract

This paper analyses the contribution of survey data, in particular various sentiment indicators, to nowcasts of quarterly euro area GDP. The empirical application makes use of a genuine real-time dataset that is constructed from original press releases and transforms the actual dataflow into an interpretable flow of news. The news is defined in our particular example as the difference between the released values and the prediction of a mixed-frequency dynamic factor model.

Our purpose is twofold. First, we aim to quantify the specific value added for nowcasting a set of heterogeneous data releases including not only sentiment indicators constructed by Eurostat, Markit, the National Bank of Belgium, IFO, ZEW, GfK or Sentix, but also hard data regarding industrial production or retail sales in the euro area and in some of the largest euro area countries. Second, our quantitative analysis is used to define an overall ranking for the data releases. The resulting order of each indicator is determined by the Kalman filter gain and depends on both its timeliness and its predictive character for GDP.

Among the survey indicators, we find the strongest impact for the Markit Manufacturing PMI and the Business Climate Indicator in the euro area, and the IFO Business Climate and IFO Expectations in Germany. Regarding the widely monitored consumer confidence indicators included in our analysis, none of them leads to significant revisions of the nowcast. In general, hard data contribute less to the nowcasts: they may be more closely correlated with GDP but their relatively late availability implies that they can to a large extent be anticipated by nowcasting on the basis of survey data and, hence, their 'news' component is smaller. Nevertheless, industrial production for the euro area holds the fifth position in the overall ranking. Finally, we also show that the NBB's own business confidence indicator is useful to predict euro area GDP, confirming previous results. The prevalence of survey data remains also under the counterfactual scenario in which hard data are released without any delay. This result confirms that, in addition to being available in a more timely manner, survey data also contain relevant information that is not captured by hard data.

Key words: JDemetra+Nowcasting, surveys, news, dynamic factor models, press releases, real-time data, Bloomberg, Forex Factory, Kalman gain
JEL Classification: C32 , C55 ,C53, C87 .

[*]National Bank of Belgium, Economics and Research Department, e-mail: raisa.basselier@nbb.be.
[†]National Bank of Belgium, R&D Statistics, e-mail: david.deantonioliedo@nbb.be.
[‡]National Bank of Belgium, Economics and Research Department, e-mail: geert.langenus@nbb.be.

Contents

1 Introduction

nowcasting and the use of kalmal filtering and factor models to was developed in my Jme paper, based on the two step approach. Peres Quiros and Banbura and Modugno did it with maximum likelihood, which was also done in the handbook chapters of Banbura et al, 2011 and 2013.

In the field of economics, the term *nowcasting* generally refers to methods for forecasting the current state of the economy and developments in the short term, and it has become increasingly popular since Giannone, Reichlin and Small (2008) and Evans (2005). After all, it is definitely relevant for both policy makers and market participants to be informed about the real activity growth in a timely manner before taking certain decisions, but official national accounts data only come with a substantial delay. Up until recently, Eurostat released the gross domestic product (GDP) 'flash' figure for the aggregate euro area only with an approximate delay of 45 days.[1] Luckily, certain higher-frequency variables, either 'hard' or 'soft' information, are published at an earlier stage and journalists, analysts in public institutions and financial market participants can already form their expectations on the basis of this dataflow.

Our main goal is to convert the heterogeneous dataflow that enters the forecaster's information set into a newsflow that can be interpreted and, most importantly, quantified. This allows us to determine the contribution of each piece of news on the GDP forecast revisions over the whole forecasting path. As stressed by Bańbura et al. (2011), the practice of nowcasting should go beyond the simple production of an early estimate and also requires the assessment of the impact of new data on forecasting updates over the time horizon. Indeed, market participants will continuously react to data releases in function of whether those are above or below their expectations, also depending on the perceived quality of the underlying data. In this paper we formalize the key features of such behavior as in Giannone et al. (2008) by using a dynamic factor model (DFM) in combination with Kalman filtering techniques. Models of this sort are successful at capturing the business cycle comovements in terms of few underlying factors and have

[1] As of 29 April 2016, Eurostat publishes a *preliminary* flash GDP with a delay of about 30 days. A second GDP flash estimate will continue to be published about 45 days after the end of the reference quarter. In this paper, we will not take on board the preliminary flash publication, as the time span covered is too limited.

been applied in many countries.[2] The advantage of the mixed-frequencies DFM proposed in this paper is the that a multivariate representation of all the data can be estimated with maximum likelihood, as Bańbura and Modugno (2010) or Camacho and Pérez Quirós (2010). Our analysis of contributions will be obtained in function of the Kalman filter gain, which can be easily applied in the context of the dynamic factor model used in this paper. From the methodological point of view, however, our proposal is not restricted to DFMs, but it simply requires a fully specified dynamic system that can be written in state-space form. Hence, linking news and forecasting updates is impossible in the context of partial models such as bridge equations, MIDAS regressions and univariate models in general, which remain widely used in nowcasting applications.

Second, the incremental contributions of the news resulting from our empirical results allow us to create a ranking for the multiple press releases. Such a ranking, which is determined by both the model parameters and the calendar of data releases, could provide a powerful tool for the aforementioned analysts and observers that are keeping track of the newsflow on a regular basis. The ranking would allow them to filter the huge amount of incoming information and mainly focus on the most important indicators. From a more formal point of view, we will investigate whether such ranking could be used as a variable selection criterion.

The empirical application in this paper exploits information from a genuine real-time dataset. We are the first ones to use time series constructed on the basis of original press releases, and our data covers a period of fifteen years. We stress the fact that modelling news and calculating the impact on forecasting revisions requires dealing with two key characteristics of real-time data. First, the dataset will have a *ragged-edge*, reflecting data availability according to the release schedule. This way of organizing the data has become standard practice in nowcasting applications since Giannone et al. (2008). Some examples based on euro area data are offered by Angelini et al. (2011) or Gayer et al. (2015), who conduct forecasting evaluation exercises compatible with the real-time publication schedule, but ignore the presence of data revisions. This is often referred to as the *pseudo real-time* approach.

The second feature of the dataset is more subtle, because we also aim to avoid distortions related with *revisions* in the indicators while respecting their release schedule. Our

[2]Please refer to Bańbura, Giannone, Modugno and Reichlin (2013) for an overview.

proposal to build a time series of first releases is not conventional, but it responds to the need of quantifying the precise impact of the first estimates of the various indicators in the forecasts of a given target e.g. euro area flash GDP, which remains the main goal of this paper.

Our investigation of the role of qualitative surveys data when forecasting macroeconomic variables provides specific weights to all indicators included in the analysis. Thus, our work adds value to the results by Abberger (2007), Claveria et al. (2007), Giannone et al. (2009), Lui et al. (2011), Martinsen et al. (2014), Piette and Langenus (2014), de Antonio Liedo (2015) and Gayer et al. (2015), to name a few. However, we are the first ones to quantify a *direct* measure of surveys' informative content. Gayer et al. (2015), for example, provide an *indirect* assessment of the usefulness of a whole block of survey data by quantifying the forecasting accuracy losses resulting from a model that excludes such block. In turn, we use a unique model to determine the exact contribution of each indicator. This represents an elegant way to answer a complex question, which cannot be found in the results of Gayer et al. (2015) or other papers, which never promised to *quantify* the average impact of each predictor variable in the forecasts.

Going beyond the precise quantification of each indicator's contribution on the forecast, this paper also explores whether the incremental impacts defining our ranking could be used as a variable selection criterion, as Rünstler (2016) proposes in his alternative analysis of contributions. Interestingly, we find that the dynamic factor model containing only the highest-ranked indicators produces less accurate forecasts than the benchmark model. Hence, we argue that relying on a limited set of indicators may prove to be somewhat more risky in the real-time environment, as opposed to the conclusion reached by Rünstler (2016), who exploits the bridging with factors framework popularized by Giannone et al. (2008).

Finally, we show that our workhorse dynamic factor model based on both hard and soft data produces euro area growth nowcasts that improve over time in function of the news distilled by the model. Those forecasting updates are not significantly different from the ones produced by the market. This implies that the parametric assumptions incorporated in our news-reading machinery are not at odds with reality, and our results can be safely regarded.

The paper is structured as follows. Section 2 describes in detail the real-time dataflow that is relevant for monitoring the business cycle in the euro area, and defines the *standard impact* concept that will be used to rank the different indicators. Section 3 presents a dynamic factor model that is flexible enough to account for a substantial proportion of the dynamic interactions between all indicators and discusses the standard impacts on the euro area GDP nowcasts that result from this model. Those results allow us to construct our ranking for the data releases. Section 4 reproduces the previous results when the target variable is German GDP but also shows the outcome of different counterfactual scenarios. These counterfactual exercises allow to disentangle the impact of the timeliness of the indicator, as well as of its quality. The last section concludes. The results presented in this paper can be easily reproduced and extended by installing a nowcasting plug-in in the JDemetra+ software, which was developed at the National Bank of Belgium.[3]

[3] JDemetra+ is free and open source software written in Java. Download it here: `https://github.com/jdemetra/jdemetra-app/releases/tag/v2.1.0-rc2`
The Nowcasting plugin should be downloaded here: `https://github.com/nbbrd/jdemetra-nowcasting/releases`. After downloading it, go to the Tools option in JDemetra+ and select *plugins*. The sofware is portable and it could even be executed from a USB disc.

2 Dataset and Analysis of News

Giannone et al. (2008) introduced the real-time dataflow as an essential element in nowcasting at a time when the literature on real-time analysis of business cycles focused mainly on the problem of data revisions while paying little attention to the fact that data sets are unbalanced at the end of the sample. Bańbura et al. (2011, 2013) emphasize that nowcasting has been taken to a whole new level by defining a mapping from surprises in new data releases onto forecasting revisions. In the first sub-section, we describe the release calendar for the euro area or the data publication process, and we do not make any assumption yet about the data generating process.

2.1 Dataset and the Real-Time Dataflow

All data used in this paper were taken from the first original press releases and therefore constitute a genuine real-time dataset. We only take on board the new figures that are provided by each publication and disregard any revisions to earlier figures. The dataset includes both soft data (survey variables) and hard data (industrial production and retail sales) for a selection of euro area countries, as shown in Table 1. The table contains an overview of the variables' definition, their frequency, publication lags and the start of the sample.

The data selection method is very simple. The dataset contains 34 monthly and quarterly series that satisfy one condition: they have been regularly distributed through the economic calendar of either Bloomberg or Forex Factory, and hence, can be considered as market moving indicators that are relevant for observers. All figures, including GDP flash and hard data, correspond to the first releases of the statistical agencies and, hence, are not affected by subsequent revisions.

There are two additional considerations in the construction of this dataset. First, all series are taken exactly as they were distributed through the Bloomberg platform. Hard data such as industrial production, retail sales or factory orders, have been distributed by Bloomberg in terms of both year-on-year growth rates and seasonally adjusted month-on-month rates. Due to the presence of data revisions, both transformations provide us with complementary information.[4] In turn, surveys have been distributed in terms of

[4]In case revisions have occurred to past data, the year-on-year growth rate is likely to give some

seasonally-adjusted balances, without further transformations.

Second, the time span has been extended using the official sources when it was deemed appropriate. As shown in Table 1, we have extended six survey indicators backwards up to 2000, or 2003 in the case of investor confidence for the euro area. We have also extended year-on-year German industrial production, which was only available since 2013. We combined Destatis historical series with the first estimates of the month-on-month growth rate to obtain a realistic estimate of the year-on-year growth rates that were realized in real-time. More specifically, we assume that after four months, the monthly growth rates are not subject to significant revisions. We argue that such an extension of our sample does not violate the conditions for a genuine real-time experiment, especially in the case of surveys, because those series have not been subject to revisions, i.e. the historical data that were used to extend the sample fully coincide with the official first releases.

For the GDP variables, we follow Kishor and Koenig (2012) and distinguish the first release from subsequent ones, but we do not make any assumption regarding the nature of data revisions. Our innovative real-time approach differs from earlier papers that exploit conventional panels of real-time data *à la* Croushore and Stark (2002). Their real-time panel for a given variable consists of increasingly large time-series in each column, which is associated to the date in which that series has been made available. Hence, the last column available, which is the one used by a forecaster today, would contain the most recent figures along with past values that have already been subject to revisions. Diron (2008), Camacho and Pérez Quirós (2010) are among the first ones to follow this approach for euro area data[5].

information on this as well.

[5] Interestingly, Camacho and Pérez Quirós (2010) do acknowledge that there is a problem with the use of vintages. For only one of the variables, GDP, they build a time series composed of flash estimates, and a different time series containing revised values. Such refinement, however, is not considered for the remaining time series. De Antonio Liedo (2015), who places more emphasis on the Belgian economy than on the euro area, follows the same approach, but also fails to reconstruct the monthly releases of hard data.

Table 1: Dataset

	Indicator	Country or Region	Source	Available on Bloomberg	Available on Forex Factory	Definition (For survey variables: Level of an index based on...)	More information	Start of our sample	Sample extension	Frequency	Released
soft data	Business climate indicator	Euro area	EC	✓	×	surveyed manufacturers, builders, wholesalers, and retailers	Survey of businesses in euro area countries	May 2001		M	around 3 weeks into the current month
	Consumer confidence	Euro area	EC	✓	✓	surveyed consumers	Survey of about 2,300 consumers in euro area countries that asks respondents to rate the relative level of past and future economic conditions, including personal financial situation, employment, inflation, and climate for major purchases	May 2001		M	around 22 days into the current month
	Economic sentiment indicator	Euro area	EC	✓	×	combined surveys	The economic sentiment indicator is composed of the industrial confidence indicator (40%), the service confidence indicator (30%), the consumer confidence indicator (20%), the construction confidence indicator (5%), and the retail trade confidence indicator (5%). Its long term average (1990-2015) equals 100.	May 2001		M	around 3 weeks into the current month
	Manufacturing PMI	Euro area	Markit	✓	✓	surveyed purchasing managers in the manufacturing industry	Survey of about 3000 purchasing managers that asks respondents to rate the relative level of business conditions including employment, production, new orders, prices, supplier deliveries, and inventories	December 2004	January 2000	M	around 3 weeks into the current month
	Investor confidence	Euro area	Sentix	✓	✓	surveyed investors and analysts	Survey of about 2,800 investors and analysts that asks respondents to rate the relative 6-month economic outlook for the Eurozone	January 2007	January 2003	M	on the first or second Monday of the current month
	Economic sentiment	Euro area	ZEW	✓	✓	surveyed German investors and analysts	Survey of up to 350 German institutional investors and analysts that asks respondents to rate the relative 6-month economic outlook for the Eurozone	November 2003	January 2000	M	on the second or third Tuesday of the current month
	Business confidence	Belgium	NBB	✓	✓	surveyed manufacturers, builders, services and trade-related firms	Survey of about 6,000 businesses that asks respondents to rate the relative level of current business conditions and expectations for the next 3 months	April 2001		M	around 3 weeks into the current month
	Consumer confidence	Belgium	NBB	✓	×	surveyed consumers	Survey of about 1850 households on their current appreciation and expectations for the next 12 months on the outlook for the general economy and regarding their own situation	January 2004	January 2000	M	around 19 days into the current month
	Consumer confidence	Germany	GfK	✓	✓	idem	Survey of about 2,000 consumers that asks respondents to rate the relative level of past and future economic conditions, including personal financial situation, climate for major purchases, and overall economic situation	January 2005		M	around the end of the previous month
	Business climate indicator	Germany	IFO	✓	✓	surveyed manufacturers, builders, wholesalers, and retailers	Survey of about 7,000 businesses that asks respondents to rate the relative level of current business conditions and expectations for the next 6 months	April 2001		M	around 3 weeks into the current month
	Business expectations	Germany	IFO	✓	×	idem		February 2002		M	idem
	Current assessment	Germany	ZEW	✓	×	surveyed German institutional investors and analysts	Survey of up to 350 German institutional investors and analysts that asks respondents about the general state of the economy as it relates to businesses	February 2004	January 2000	M	on the second or third Tuesday of the current month
	Economic sentiment	Germany	ZEW	✓	✓	idem	Survey of up to 350 German institutional investors and analysts that asks respondents to rate the relative 6-month economic outlook for Germany	December 2001	January 2000	M	idem
hard data	Industrial production m/m	Euro area	EC	✓	✓	Change in the total inflation-adjusted value of output produced by manufacturers, mines, and utilities		July 2001		M	about 45 days after the month ends
	Industrial production y/y	Euro area	EC	✓	×	idem		July 2001		M	idem
	Industrial production m/m	Germany	Destatis	✓	✓	idem		January 2006		M	about 40 days after the month ends
	Industrial production y/y	Germany	Destatis	✓	×	idem		July 2013	January 2006	M	idem
	Factory orders m/m	Germany	Destatis	✓	✓	Change in the total value of new purchase orders placed with manufacturers		April 2001		M	about 35 days after the month ends
	Factory orders y/y	Germany	Destatis	✓	×	idem		April 2001		M	idem
	Retail sales m/m	Germany	Destatis	✓	✓	Change in the total value of inflation-adjusted sales at the retail level, excluding automobiles and gas stations		June 2003		M	about 30 days after the month ends
	Retail sales y/y	Germany	Destatis	✓	×	idem		June 2003		M	idem
	Retail sales y/y	Spain	INE	✓	×	idem		June 2013		M	about 30 days after the month ends
	Retail sales m/m	Italy	Istat	✓	✓	Change in the total value of sales at the retail level		October 2003		M	about 55 days after the month ends
	Retail sales y/y	Italy	Istat	✓	×	idem		October 2003		M	idem
	Industrial production m/m	Italy	Istat	✓	✓	Change in the total inflation-adjusted value of output produced by manufacturers, mines, and utilities		May 2001		M	about 40 days after the month ends
	Industrial production y/y	Italy	Istat	✓	×	idem		May 2001		M	idem
	Industrial production m/m	France	INSEE	✓	✓	idem		April 2001		M	about 40 days after the month ends
	Industrial production y/y	France	INSEE	✓	×	idem		April 2001		M	idem
	Consumer spending m/m	France	INSEE	✓	✓	Change in the inflation-adjusted value of all goods expenditures by consumers		April 2001		M	about 27 days after the month ends
	Consumer spending y/y	France	INSEE	✓	×	idem		April 2001		M	idem
GDP	GDP flash	Belgium	NAI	✓	✓	Change in the total inflation-adjusted value of all goods and services produced by the economy		September 2003		Q	about 30 days after the quarter ends
	GDP flash forecast	Euro area	Bloomberg	✓	×	idem		June 2001		Q	about 44 days after the quarter ends
	GDP flash	Germany	Destatis	✓	✓	idem		June 2003		Q	about 45 days after the quarter ends
	GDP flash	Euro area	EC	✓	✓	idem		March 2001		Q	idem

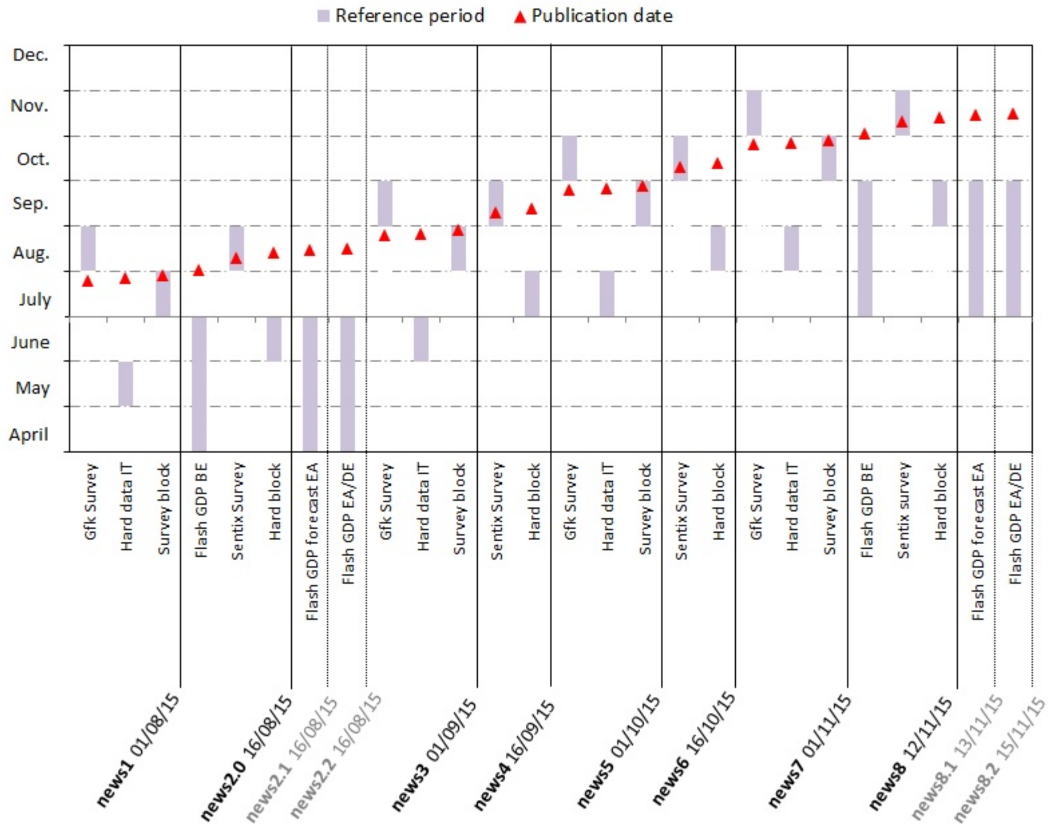
The dataset generally exhibits a ragged-edge pattern. The 34 variables are subject to different publication lags, as specified in the last column of Table 1. Moreover, two data releases distributed around the same time may refer to an entirely different reference period, especially as our dataset includes both soft and hard data, of which the latter are typically subject to longer publication lags. To make it easier to visualize the characteristics of the dataset used, a simple description of the dataflow is displayed in Figure 1, taking the August-November 2015 period as a reference. The horizontal axis represents the sequence of variables that are released. The vertical axis represents the dates at which each one of those variables are released (cf. triangles) and at the same time it signals their reference period using bars. Figure 1 is a more simple representation of Figure 2, which fully demonstrates the complexity of the problem. Although the data are released in a continuous manner, we simplify the analysis by proposing a regular updating scheme that takes place only twice per month. As argued by Angelini et al. (2011), European data releases tend to be clustered at the end or the middle of the month anyway. The vertical lines in the graphs represent dates where GDP nowcasts will be updated. The impact of the multiple variables will be calculated by *block of news*.

- The first block of news, which is *read* on 1 August 2015, considers the Gfk survey for the month of August and the remaining surveys for July, i.e. by EC, Markit, NBB, IFO and ZEW. It also contains some hard indicators (retail sales and industrial production) for Italy for the month of May. Each piece of news will contribute to updating the growth forecasts for the euro area.

- Two weeks after, the second block of data (2.0) will be read. This block contains mainly hard indicators that still refer to the month of June: industrial production, factory orders and retail sales for Germany, consumer spending and industrial production for France, as well as industrial production for the aggregate euro area. Only one survey variable is included: the Sentix release regarding uncertainty for the month of August. Apart from those monthly indicators, the release of Belgian flash GDP for the second quarter, which is typically published by the National Accounts Institute one month after the end of each quarter, also belongs to the second block of news. Whether such national releases can help to forecast the euro area aggregate is an empirical question for which the method described in this paper provides a straightforward answer.

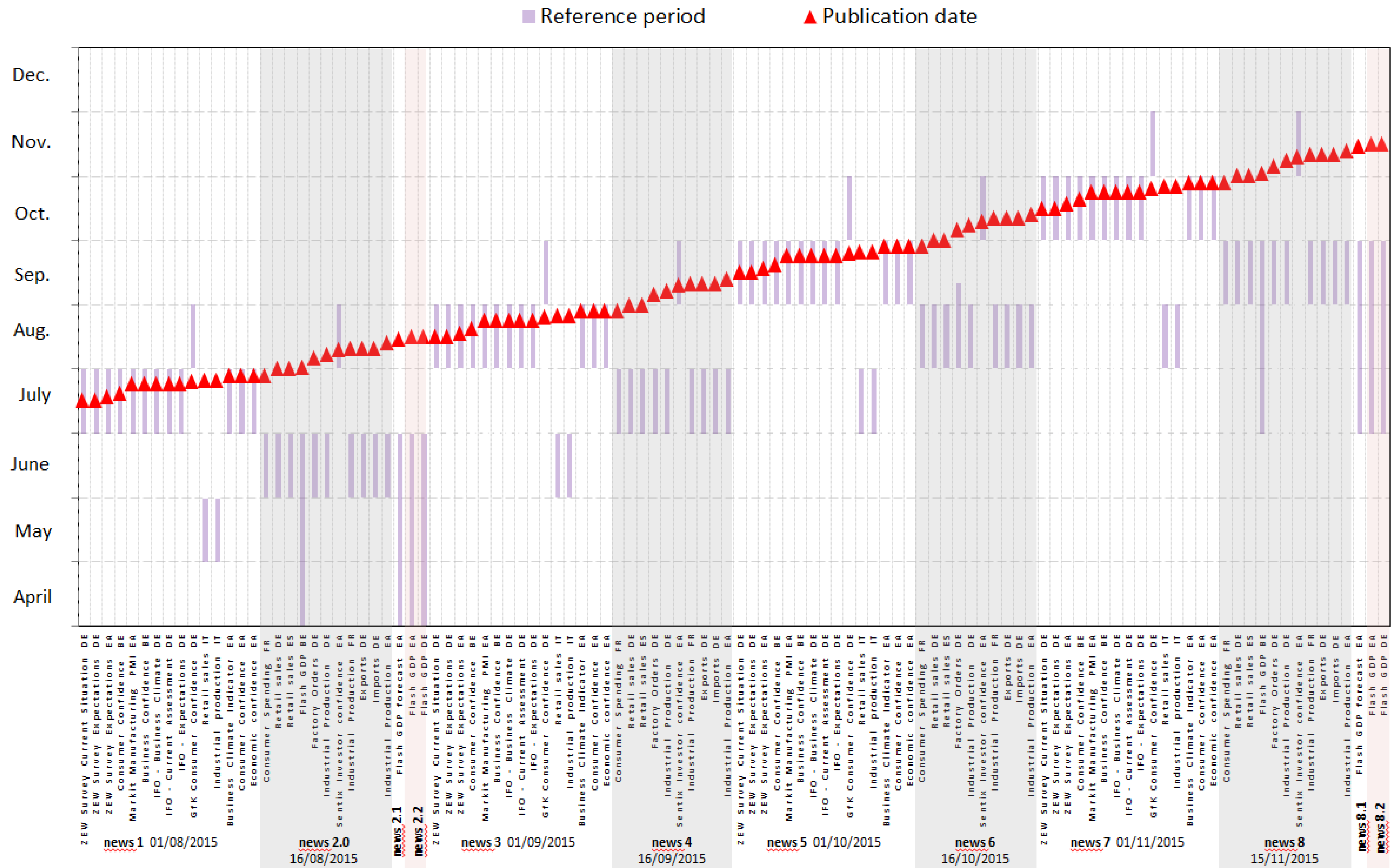
- Separately from the second block, we have the official flash estimate of real GDP growth rates for the euro area and Germany (Block 2.2), which are often read in terms of how close they are to the forecast expected by market participants. Such a forecast is also represented in the graph (Block 2.1), and we will assume it is known right before the official flash release.
- The blocks $3-4$ and $5-6$ represent the sequential arrivals of soft and hard data, while the blocks $7-8.0-8.1-8.2$ are a repetition of the previous two points.

Figure 1: Representation of the Dataflow



This figure represents the arrival of macroeconomic news that market participants are faced with. Although the data are actually released in a continuous manner, we aim to simplify the analysis by proposing a regular updating scheme that takes place only twice per month: around the middle of each month, i.e. after most hard indicators are published, and two weeks later, i.e. when most surveys have been released. Notice that the flash GDP for the euro area in this graph still refers to the flash with a publication delay of 45 days, and does not yet incorporate the recently introduced *preliminary flash release* with a delay of only 30 days. In this paper, we define the news as the forecast errors obtained for all variables when the forecasts are updated twice a month. We will also emphasize the news content of the GDP releases for the euro area and Germany, which take place 45 days after each quarter ends.

Figure 2: Detailed View of the Real-Time Dataflow



This figure represents the real-time dataflow. All data releases published between June 2015 and October 2015 are represented chronologically. Most of the surveys for each month or reference period are published before the end of the month. This means that the publication dates marked with triangles often fall inside the bar representing reference periods. In the extreme, some variables' with a strong expectations components are released prior to the reference period. Conversely, variables subject to publication lags, such as GDP or industrial production, will have the triangle way above the reference period. Hard data such as industrial production and retail sales are included in both month-on-month and year-on-year growth rates, except for Spanish retail sales, which are only included in month-on-month growth rates.

2.2 Defining the Newsflow and its Impact on Euro Area real GDP growth

This subsection clarifies the concept of news, and how the real-time dataflow is translated into a newsflow.

The *news* associated with a given release is represented by the discrepancy of the published figure with respect to the expected value. Thus, this concept depends on expectations, which in our case will be model-consistent. The word *news*, *innovations* or *forecast errors* will be used interchangeably (see Durbin and Koopman, 2001). Once the concept of news is clarified, we will show how the *Kalman gain* induces the model to update the forecast path for GDP or any other variable of interest after a given piece of news becomes available. The impact of the news that gradually enters the forecaster's information set is given not only by their quality, but also by their timeliness, which crucially depends on the release calendar. Our particular schedule for data releases and the publication lags associated to each indicator are represented in Figure 2.

Let's consider a generic recursive representation for the observable indicators described before:

$$\mathrm{y_t} = \Lambda \mathrm{f_t} + \mathrm{e_t} \tag{1}$$

$$\mathrm{f_{t+1}} = A\mathrm{f_t} + \eta_t \tag{2}$$

with normally distributed measurement errors and shocks to the factors: $e_t \sim N(0, R_t)$, $\eta_t \sim N(0, Q_t)$.

Defining the information sets

The concept of news can be formalized by specifying information sets that enter the model. Figure 3 displays the two consecutive information sets that are used to define, for example, news 8.0. In order to simplify the notation and without loss of generality, we will assume in this exposition that there are only two news components and there are no revisions.

$\mathcal{F}^{old}$ contains all time series available right before the publication of the news. Consider for the sake of simplicity that all observations are available until time t.

$\mathcal{F}^{new}$ includes the previous information set, $\mathcal{F}^{old}$, plus new data corresponding to a given macroeconomic release. Again for the sake of simplicity, one can assume that the release extends by one month, $(t+1)$, two of the indicators referring to sales $(\mathrm{y}^{\mathrm{s}}_{\mathrm{t+1}})$ and manufacturing $(\mathrm{y}^{\mathrm{m}}_{\mathrm{t+1}})$.

The forecast for the whole vector of variables y_{t+h} is formulated in our framework in terms of model consistent conditional expectations:

$$E_\theta[y_{t+h}|\mathcal{F}^{new}] = E_\theta[y_{t+h}|\mathcal{F}^{old}, \{V_{t+1}\}] \tag{3}$$

where the expression on the right-hand side decomposes the new conditioning information set in two orthogonal parts. In this particular example, $V_{t+1} \equiv \mathcal{F}^{new} - \mathcal{F}^{old} = [\mathrm{v}^{\mathrm{m}}_{\mathrm{t+1}} \;\; \mathrm{v}^{\mathrm{s}}_{\mathrm{t+1}}]'$ incorporates two innovations or news, which are defined as the difference between the released figures and the model's forecasts:

$$\mathrm{v}^{\mathrm{m}}_{\mathrm{t+1}} = \mathrm{y}^{\mathrm{m}}_{\mathrm{t+1}} - E_\theta[\mathrm{y}^{\mathrm{m}}_{\mathrm{t+1}}|\mathcal{F}^{old}]$$

$$\mathrm{v}^{\mathrm{s}}_{\mathrm{t+1}} = \mathrm{y}^{\mathrm{s}}_{\mathrm{t+1}} - E_\theta[\mathrm{y}^{\mathrm{s}}_{\mathrm{t+1}}|\mathcal{F}^{old}]$$

Thus, one could state that, even if the released figures have declined with respect to the recent past, the model could interpret them as good news as long as they are above the values that the model was expecting.

The Kalman filter gain

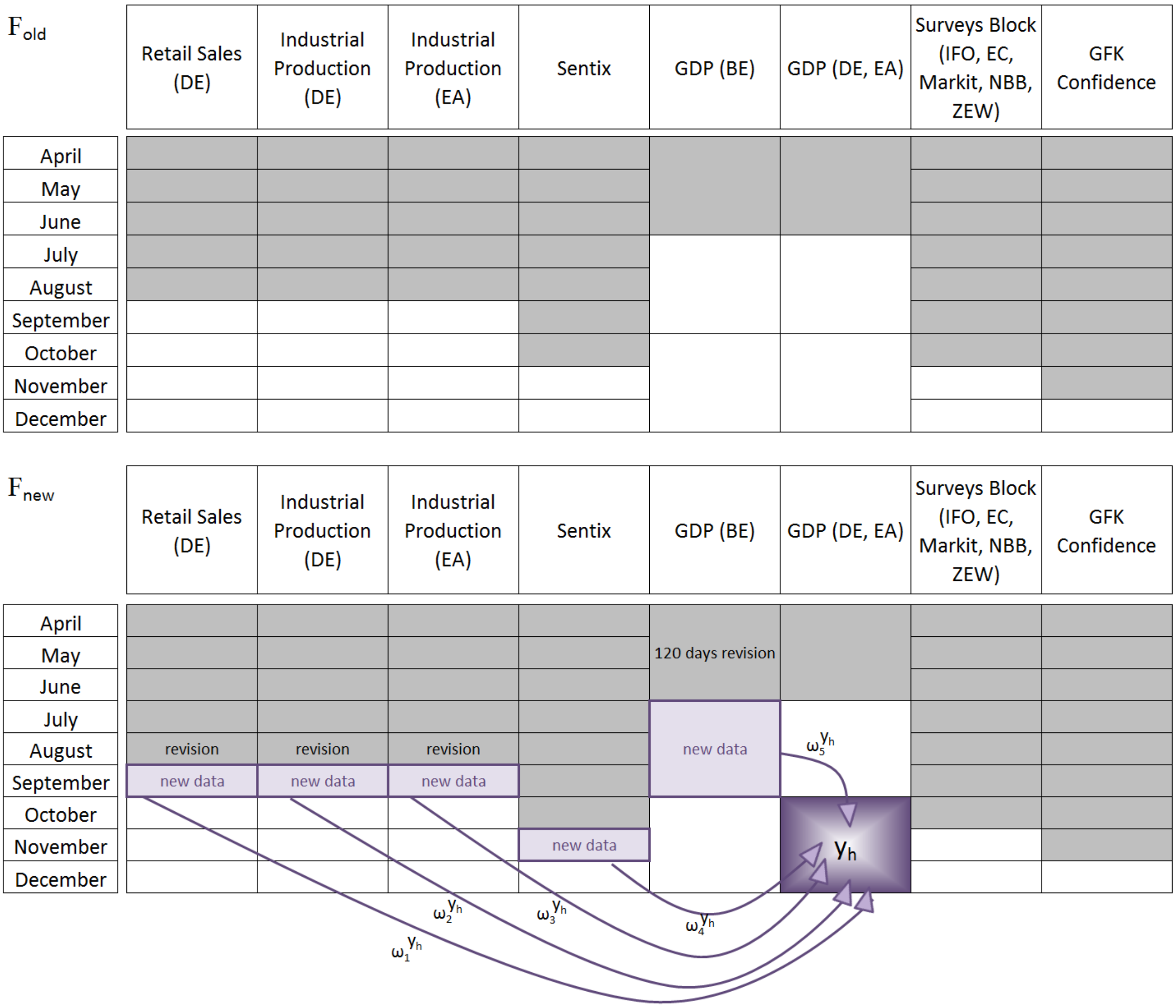
This news is exploited by the Kalman filter *gain* in order to update GDP forecasts together with the remaining variables. If we could observe $\mathrm{f}_{\mathrm{t+h}}$, obtaining the forecast would be straightforward: $E_\theta[y_{t+h}|\mathcal{F}^{new}] = \Lambda A^{h-1} f_{t+1}$. But unfortunately, the factor $\mathrm{f}_{\mathrm{t+1}}$ cannot be observed because only two data releases for $t+1$ are available and they are assumed to contain measurement errors. Thus, the conditional expectation in expression 3 must be developed further:

$$\begin{aligned} E_\theta[y_{t+h}|\left\{\mathcal{F}^{old}, V_{t+1}\right\}] &= \Lambda A^{h-1} E_\theta[\mathrm{f}_{\mathrm{t+1}}|\left\{\mathcal{F}^{old}, V_{t+1}\right\}] \\ &= \underbrace{\Lambda A^{h-1} E_\theta[\mathrm{f}_{\mathrm{t+1}}|\mathcal{F}^{old}]}_{old\,forecast} \\ &+ \Lambda A^{h-1} \underbrace{E_\theta[\mathrm{f}_{\mathrm{t+1}} V'_{t+1}] E_\theta[V_{t+1} V'_{t+1}]^{-1}}_{Gain\,(quality,\,timeliness)} \underbrace{V_{t+1}}_{news} \end{aligned} \tag{4}$$

All the expectations are calculated with the smoothed covariance, which is given by the precision of the filter. Interestingly, the product of expectations shown in the expression above defines how the *news* induces an update[6] of the state of the economy, which is represented by $\mathrm{f}_{\mathrm{t+1}}$. The precise weight of each one of the innovations at updating the expectations about the state of the economy depends on the quality and the timeliness of the indicators. By quality, we refer to the correlation of the factor with the innovations $E_\theta[\mathrm{f}_{\mathrm{t+1}} V'_{t+1}]$. The role of timeliness, which determines the pattern of missing observations, is also crucial at defining the weights. Thus, it can be easily understood that variables that enter first the model's information set will receive a larger weight than in the case where they are part of a larger group of data releases. The reason is that in the presence of strong collinearity where all variables incorporate the same signal, only one variable is enough.

[6]This update takes the form of a simple OLS regression of the factors on the innovations. Note that the size of the news vector V_{t+1} may be large in practical applications. For example when many variables can be released at the same time, or many observations for the same variable are made available simultaneously.

Figure 3: *News 8.0* Block's weights for updating euro area GDP for next quarter



This figure represents two consecutive information sets and the weights associated to the news represented in Figure 1 as 8.0. Note that the weight's subindex corresponds to each one of the pieces of news, while the upperindex refers to the target variable, which is next quarter's GDP growth (y_h). Thus, the weight of those news at updating current quarter's GDP would be represented with the superindex y_{h-1}. Although press releases may contain revisions for certain (hard) indicators, these will not be taken on board in our empirical application as we are only interested in the first vintages (i.e. new data).

Defining the standard impact of news

In this simple example with only two data releases and one factor, the last term of expression 4 can be written in terms of parameters $\sigma_m^2 = var_\theta(v_{t+1}^m)$, $\sigma_s^2 = var_\theta(v_{t+1}^s)$, $\sigma_{ms}^2 = cov_\theta[v_{t+1}^m, v_{t+1}^s]$, $\beta_m = cov_\theta[\mathrm{f_{t+1}}, v_{t+1}^m]$, $\beta_s = cov_\theta[\mathrm{f_{t+1}}, v_{t+1}^s]$. Thus:

$$\begin{aligned} E_\theta[y_{t+h}|\left\{\mathcal{F}^{old}, V_{t+1}\right\}] - E_\theta[y_{t+h}|\left\{\mathcal{F}^{old}\right\}] &= \underbrace{\Lambda A^{h-1}\frac{\beta_m\sigma_s^2 - \beta_s\sigma_{ms}^2}{\sigma_m^2\sigma_s^2 - \sigma_{ms}^2\sigma_{ms}^2}}_{impact\ of\ manufacturing} v_{t+1}^m \\ &+ \underbrace{\Lambda A^{h-1}\frac{\beta_s\sigma_m^2 - \beta_m\sigma_{ms}^2}{\sigma_m^2\sigma_s^2 - \sigma_{ms}^2\sigma_{ms}^2}}_{impact\ of\ sales} v_{t+1}^s \end{aligned} \tag{5}$$

This illustration has served as a vehicle to underline that the whole set of news, i.e. the vector of innovations V_{t+1}, induces an update of the path for all variables in y_t. The extent to which all the individual pieces of news induce change expectations for GDP growth rates in the euro area depends on all the different factors and on the particularities of the calendar of data releases. Quantifying the precise role of all the news is the goal of the next section. By multiplying the impacts defined in the equations above by the standard deviation of the news associated with each data release, we obtain a measure that allows us to compare the average informative content of the different indicators when the object of interest is real economic growth, as measured by GDP.

3 Empirical Results

This section quantifies the impacts of all data releases with respect to updating the forecasts for real GDP growth rates in the euro area on the basis of the methodology explained in Section 2. We will rank all the news releases according to their expected impacts for updating real GDP growth forecasts for the euro area. Those expected impacts will be given by a parametric dynamic factor model representing all monthly and quarterly variables. Thus, the state-space form mentioned earlier will remain valid. Below, we outline the most common approach to link GDP, which is the quarterly variable that we use as a target, with the unobserved factors, which are specified at a monthly frequency and determine the joint dynamics of the whole system.

3.1 A State-Space Representation

We describe here a joint model for the German economy and the aggregate euro area data. Germany was included because it could be considered as one of the main drivers of the euro area business cycle, given its size and the high share of the manufacturing sector in total value added in Germany compared with other euro area countries (ECB, 2011). We assume that there are four underlying factors, but for the sake of simplicity, the following expression links the monthly growth rates of the variables to the vector of only two of such monthly factors:

$$\begin{aligned} y_t &= \bar{y} + \Lambda_y f_{1,t} + \Theta_y f_{2,t} + \psi_t && \textit{Measurement euro area} && (6) \\ x_t &= \bar{x} + \Lambda_x f_{1,t} + \Theta_x f_{2,t} + \chi_t && \textit{Measurement Germany} && (7) \end{aligned}$$

German time series will be denoted by x_t, while euro area series are represented by y_t. The error terms χ_t and ψ_t are assumed to be uncorrelated with the factors at all leads and lags. They are also assumed to be independent and identically distributed (iid) and following a normal distribution: $\chi_t \sim N(0, R_\chi)$ and $\psi_t \sim N(0, R_\psi)$. Both covariance matrices are assumed to be diagonal, which implies that the factors will account for 100% of the comovements implicit in the model. As suggested by Doz et al. (2012), this assumption is not very restrictive. They show that Quasi-ML estimation of the factors is consistent even in the presence of weak cross-correlation patterns in the error term.

Because the model has been designed for short-term analysis, it makes sense to represent all these series, including GDP, in terms of monthly growth rates or differences. However, the monthly growth rates of official GDP figures are not published, equations 6 and 7 need to be modified. Thus, GDP growth rates published by the statistical agencies (i.e. y_t^Q for the euro area and the x_t^Q for Germany) are linked to the quarterly growth rates of the underlying factors, which can be expressed as a moving average of their monthly growth rates:

$$\begin{aligned} y_t^Q &= \bar{y}^Q + \Lambda_y^Q f_{1,t}^Q + \Theta_y^Q f_{2,t}^Q + \psi_t^Q, \qquad && t = 3, 6, 9, \ldots \quad \textit{euro area GDP} \qquad (8) \\ x_t^Q &= \bar{x}^Q + \Lambda_x^Q f_{1,t}^Q + \Theta_x^Q f_{2,t}^Q + \chi_t^Q, \qquad && t = 3, 6, 9, \ldots \quad \textit{German GDP} \qquad (9) \end{aligned}$$

where

$$\begin{aligned} f_{1,t}^Q &= \frac{1}{3} f_{1,t} + \frac{2}{3} f_{1,t-1} + f_{1,t-2} + \frac{2}{3} f_{1,t-3} + \frac{1}{3} f_{1,t-4} \\ f_{2,t}^Q &= \frac{1}{3} f_{2,t} + \frac{2}{3} f_{2,t-1} + f_{2,t-2} + \frac{2}{3} f_{2,t-3} + \frac{1}{3} f_{2,t-4} \end{aligned}$$

As mentioned above, $f_{1,t}$ and $f_{2,t}$ represent monthly growth rates of the latent factors. The last expressions for $f_{1,t}^Q$ and $f_{2,t}^Q$ are based on the technical assumption that the quarterly level of the factors can be represented by the geometric mean of the latent monthly levels.[7] This assumption makes it possible to obtain a simple expression for the quarterly growth rate of the factors as a moving average of the latent monthly growth rates. Because we apply the Mariano and Murasawa (2003) approximation to the factors alone, and not to the observables, the error terms χ_t^Q and ψ_t^Q are assumed to be iid normally distributed and uncorrelated with all factors at all leads and lags.

So far, we have described the measurement equation, which defines the link between the unobserved factors and the two types of observable time series: monthly variables and

[7] The approximation proposed by Mariano and Murasawa (2003) is applied to the factors. Let F_t be the monthly *level* of the economy and let $f_t = lnF_t - lnF_{t-1}$ be its monthly growth rate. Now, define F_t^Q as the geometric mean of the last three levels. This implies that $lnF_t^Q = \frac{1}{3}(lnF_t + lnF_{t-1} + lnF_{t-2})$. The resulting quarterly growth rate of the factors, which we denote as f_t^Q, can be expressed as $lnF_t^Q - lnF_{t-3}^Q$. By substituting both terms by the geometric mean approximation we obtain $f_t^Q = \frac{1}{3}(lnF_t - lnF_{t-3}) + \frac{1}{3}(lnF_{t-1} - lnF_{t-4}) + \frac{1}{3}(lnF_{t-2} - lnF_{t-5})$. Finally, a simple expression for the quarterly growth rate of the factors in terms of their monthly growth rates can be obtained as follows: $f_t^Q = \frac{1}{3}(f_t + f_{t-1} + f_{t-2}) + \frac{1}{3}(f_{t-1} + f_{t-2} + f_{t-3}) + \frac{1}{3}(f_{t-2} + f_{t-3} + f_{t-4})$. Rearranging terms yields the expression $f_t^Q = \frac{1}{3} f_t + \frac{2}{3} f_{t-1} + f_{t-2} + \frac{2}{3} f_{t-3} + \frac{1}{3} f_{t-4}$ presented above.

quarterly variables (e.g. GDP). Specifying the joint dynamics of all variables in both the euro area and Germany requires a second equation representing the factors as a vector autoregressive (VAR) process with a non-diagonal covariance matrix for the error term. To sum up, representation given by equations 10 and 11 conforms to the so-called state-space representation of this model and determines the joint dynamics of both the euro area and the German business cycle:

$$\left(\begin{array}{c} x_t^Q - \bar{x}^Q \\ x_t - \bar{x} \\ \hline y_t^Q - \bar{y}^Q \\ y_t - \bar{y} \end{array}\right) = \left(\begin{array}{ccccc|ccccc} \Lambda_x^Q & 2\Lambda_x^Q & 3\Lambda_x^Q & 2\Lambda_x^Q & \Lambda_x^Q & \Theta_x^Q & 2\Theta_x^Q & 3\Theta_x^Q & 2\Theta_x^Q & \Theta_x^Q \\ \Lambda_x & 0 & 0 & 0 & 0 & \Theta_x & 0 & 0 & 0 & 0 \\ \hline \Lambda_y^Q & 2\Lambda_y^Q & 3\Lambda_y^Q & 2\Lambda_y^Q & \Lambda_y^Q & \Theta_y^Q & 2\Theta_y^Q & 3\Theta_y^Q & 2\Theta_y^Q & \Theta_y^Q \\ \Lambda_y & 0 & 0 & 0 & 0 & \Theta_y & 0 & 0 & 0 & 0 \end{array}\right) \left(\begin{array}{c} f_{1,t} \\ f_{1,t-1} \\ f_{1,t-2} \\ f_{1,t-3} \\ f_{1,t-4} \\ \hline f_{2,t} \\ f_{2,t-1} \\ f_{2,t-2} \\ f_{2,t-3} \\ f_{2,t-4} \end{array}\right) + \left(\begin{array}{c} \chi_t^Q \\ \chi_t \\ \hline \psi_t^Q \\ \psi_t \end{array}\right) \quad (10)$$

$$\left(\begin{array}{c} f_{1,t} \\ f_{1,t-1} \\ f_{1,t-2} \\ f_{1,t-3} \\ f_{1,t-4} \\ \hline f_{2,t} \\ f_{2,t-1} \\ f_{2,t-2} \\ f_{2,t-3} \\ f_{2,t-4} \end{array}\right) = \left(\begin{array}{ccccc|ccccc} A_{11} & 0 & 0 & 0 & 0 & A_{12} & 0 & 0 & 0 & 0 \\ I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline A_{21} & 0 & 0 & 0 & 0 & A_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \end{array}\right) \left(\begin{array}{c} f_{1,t-1} \\ f_{1,t-2} \\ f_{1,t-3} \\ f_{1,t-4} \\ f_{1,t-5} \\ \hline f_{2,t-1} \\ f_{2,t-2} \\ f_{2,t-3} \\ f_{2,t-4} \\ f_{2,t-5} \end{array}\right) + \left(\begin{array}{c} u_t^f \\ 0 \\ 0 \\ 0 \\ 0 \\ u_t^g \\ 0 \\ 0 \\ 0 \\ 0 \end{array}\right) \quad (11)$$

where the innovations to the factors are allowed to be cross-correlated:

$$\left(\begin{array}{c} u_t^f \\ u_t^g \end{array}\right) \sim \mathcal{N}\left(0, \left[\begin{array}{cc} Q_1 & \Omega_{12} \\ \Omega_{12} & Q_2 \end{array}\right]\right)$$

These error components are also uncorrelated with all measurement error terms, in line with the literature on factor models. For simplicity, and in contrast to the model built by Mariano and Murasawa (2003), we do not incorporate autocorrelation in the measurement errors. This helps to keep the size of the state vector as small as possible without restricting the extent to which the factors can account for the business cycle comovements.

3.2 Estimation in the Context of Missing Observations

Once the building blocks of the model have been defined, we need to tackle the problem of estimation. The Quasi-maximum likelihood procedure of Doz et al. (2012) is used here with the aim of achieving a consistent estimation even in the presence of weak correlation patterns in the measurement errors. Thus, the model is estimated under the restriction that the off-diagonal elements of the measurement error covariance matrix are equal to zero. This has the practical implication that one hundred percent of the cross-correlation patterns generated by the model will be fully accounted for by the factors.

The model is estimated at monthly frequency with maximum-likelihood even in the presence of missing observations. For example, survey data for the euro area are often not available prior to 2000, while some of the Belgian series date back to 1980. The presence of quarterly data also generates additional missing observations, since they are treated as indicators that are observed every third month of the quarter, i.e. y_t^Q as a missing variable for $t \neq 3, 6, \ldots$. Finally, as in most macroeconomic forecasting applications, the relevant information set is based on indicators that arrive gradually throughout the quarter and with important delays with respect to the period of time to which they refer, i.e. the real-time dataflow. Thus, in practice, it is unavoidable to have missing values at the end of the sample. For a detailed overview of the estimation method used in this paper, the reader is referred to Bańbura and Modugno (2010). Below, we summarize the most important concepts underlying the approach with special emphasis on the aspects that are particularly relevant in our nowcasting framework:

- **Maximum-likelihood**. In this application, the state-space model represented by equations 10 and 11 is estimated with the Expectation-Maximization (EM) algorithm. The Kalman (1960) filter and smoothing recursions, however, need to be slightly modified so that only the actual observations can be taken into account in the estimation of the factors and the evaluation of the likelihood. The EM algorithm was derived by Shumway and Stoffer (1982) only for the case where the factor loadings multiplying the factors in the measurement equation are known. Bańbura and Modugno (2010) are the first ones to apply this algorithm to the current set-up, where the loadings need to be estimated in the context of missing observations. They show that their method is consistent and computationally feasible even in the case where the number of variables is large. Alternatively, Camacho and Perez-Quiros (2010) propose the use of standard optimization routines to maximize the likelihood of a model of the same class, but based on a smaller number of variables.[8]

- **Identification of the factors**. The strongest assumption, which is key for identification, is that the measurement errors in expression 10 are uncorrelated with the factor innovations in the transition equation 11. This allows for a clear-cut separation of the measurement errors and the signal provided by the factors. In the absence of the restrictions we impose in the factor loadings, the model would be identified only up to an invertible linear transformation. That is, applying the following transformation, $g_t^M = G f_t^M$, the transition equation $g_t^M = GA_1G^{-1}g_{t-1}^M + \ldots + GA_pG^{-1}g_{t-p}^M + Gu_t$ would be observationally equivalent to the one represented by equation 11. Nevertheless, Dempster, Laird and Rubien (1977) suggest that the EM algorithm is not affected by this lack of identification. The space generated by the factors, and thereby the projections on such space, are unaffected by the choice of G. This identification issue is well known in factor analysis and does not distort any of the results presented in this paper.

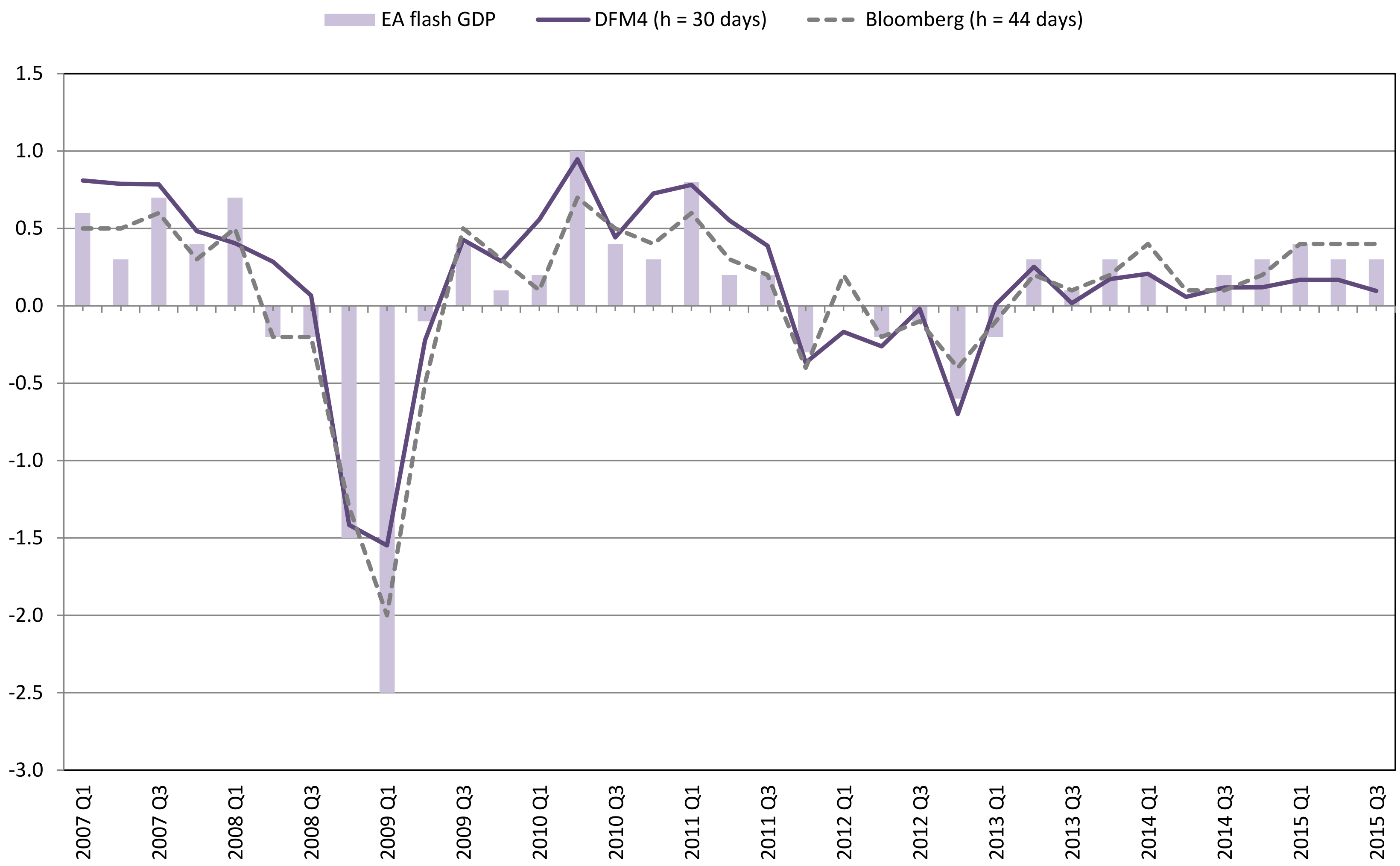
3.3 Forecasting accuracy in real-time

Before moving on to discussing the impacts that correspond to the news associated with the indicators, it is useful to make sure that the chosen model is at least able to provide a realistic representation of the news for the whole set of indicators. As a first indication, Figure 4 shows the target series (euro area flash GDP), compared to the dynamic factor model's nowcast at a horizon of 30 days after the end of the reference quarter (i.e. about two weeks before the actual flash is published by Eurostat). The nowcast provided by Bloomberg, which is usually available right before the flash publication, is also plotted in this graph.

Our analysis of news is based on a state-space representation with 4 factors. Thus, our model has a larger stochastic dimension than the Euro-Sting model developed by

[8] Numerical optimization of the likelihood, which is feasible for parsimonious models, has the advantage that it does not require the Kalman smoother. Moreover, the multithreading ability of most software packages is able to reduce the execution time by exploiting multiple processors. For example, the current estimation of dynamic factor models in *J*Demetra+ is feasible without the need of applying the EM algorithm.

Figure 4: Euro area flash GDP and the model nowcasts compared to the Bloomberg forecast



Camacho and Pérez-Quirós (2010) or the model developed by Aruoba, Diebod and Scotti (2009) for the US, which relies on a single factor and which performance crucially depends on the variables that were selected. Figure 5 shows that the model with four factors and three lags delivers the best results in terms of the root mean squared error for the euro area flash GDP. Furthermore, the graph clearly shows the decreasing pattern of the root mean squared error (RMSE) for real flash GDP growth in the euro area over the period 2006Q1-2015Q1 as more information becomes available. The error is based on forecasts obtained by re-estimating the model once a year and updating the forecasts with every data release.

As one can see in Figure 5, the RMSE associated to updates that take place with every data release (solid line) decreases gradually until the arrival of the last piece of news corresponding to each block. In practice, the strategy of exploiting in real-time the newsflow corresponding to news blocks 4 and 6, containing releases of industrial production, sales, factory orders, consumer spending and the Sentix index, has turned out to be clearly better than waiting for the last data release of each block before updating the model.

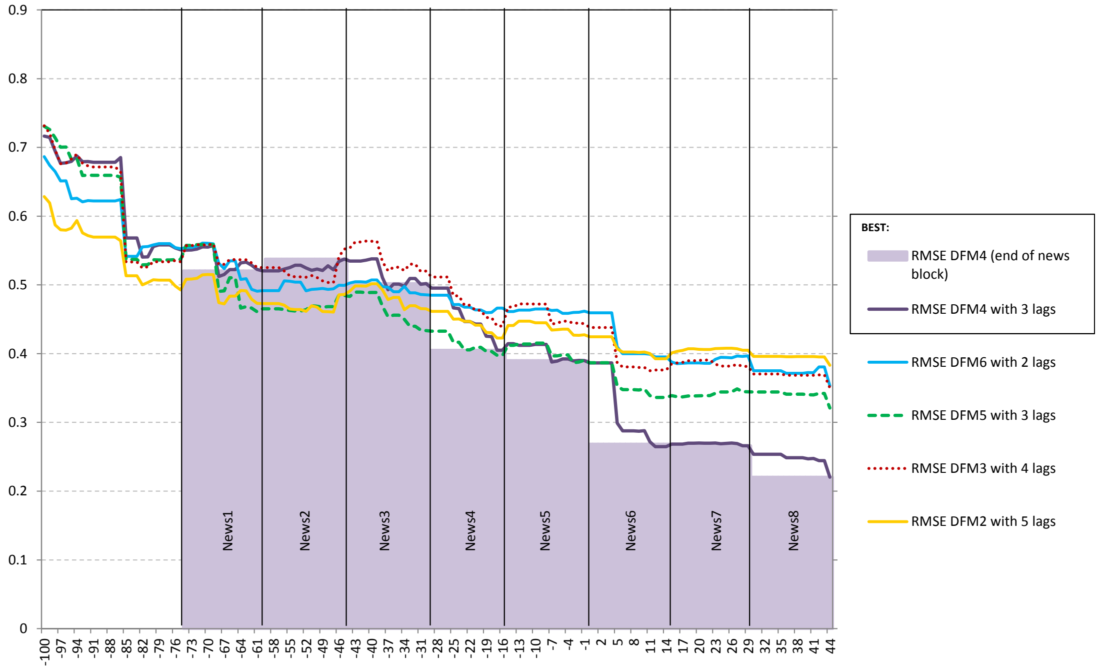
The nowcasts produced two weeks after the end of the quarter, which includes the whole block of news 6, have reduced the RMSE by 0.13. On aggregate, all the news represented in the Figure has ended up reducing the RMSE by 0.34, up to 0.22. The graph also reveals further accuracy gains over the evaluation sample when Bloomberg forecasts, which represent the only quantitative survey used in this paper, are processed by the model (cf. the small dip in the RMSE at horizon 44).

The particularities of the evaluation sample, which include the Great Recession and the sovereign debt crisis, may not be representative of a typical business cycle with long expansion periods with stable growth rates, so the 0.22 figure may be considered as an upper-bound of the forecasting uncertainty that can be expected from the model.

3.4 Ranking based on the "Standard Impact" of Macroeconomic Releases

We have defined the *standard impacts* associated with each one of the multiple news releases as the product of the impact coefficients defined in equation 5 and the standard

Figure 5: RMSE over 2006Q1-2015Q1 for different models and as a function of the expanding information set



deviation of the respective innovations, i.e. the root mean square error (RMSE) associated with the release of each series. The flow of information within the quarter has been clarified by Figures 1 and 2 in previous section. As discussed above, timely indicators will tend to have a higher weight, so it is important to take into account the calendar of data releases.

The resulting standard impacts for all data releases are displayed in Figure 6. The standard impacts are a function of the real-time dataflow. Thus, the function is *constant over time* as long as the order of the blocks data releases remain unchanged. The graph shows that some indicators consistently have a substantial impact on the revision of real GDP growth expectations in the euro area in the current quarter, or even the next. This is for example the case for the Markit PMI releases for the euro area, the German IFO surveys, the business climate indicator for the euro area and also industrial production. Other indicators, such as the the retail sales indicators or the headline consumer confidence indicators always show low impacts.[9] Piette and Langenus (2014) have already warned that restricting the dataset to headline survey indicators implies a loss of information. They find that certain sub-components of the headline confidence indicator (in particular

[9]Oddly, the impact of the GfK consumer confidence is even diving into negative territory. We believe that this result should rather be interpreted as a 'zero' impact.

the unemployment expectations) are more relevant for the early estimates of GDP growth.

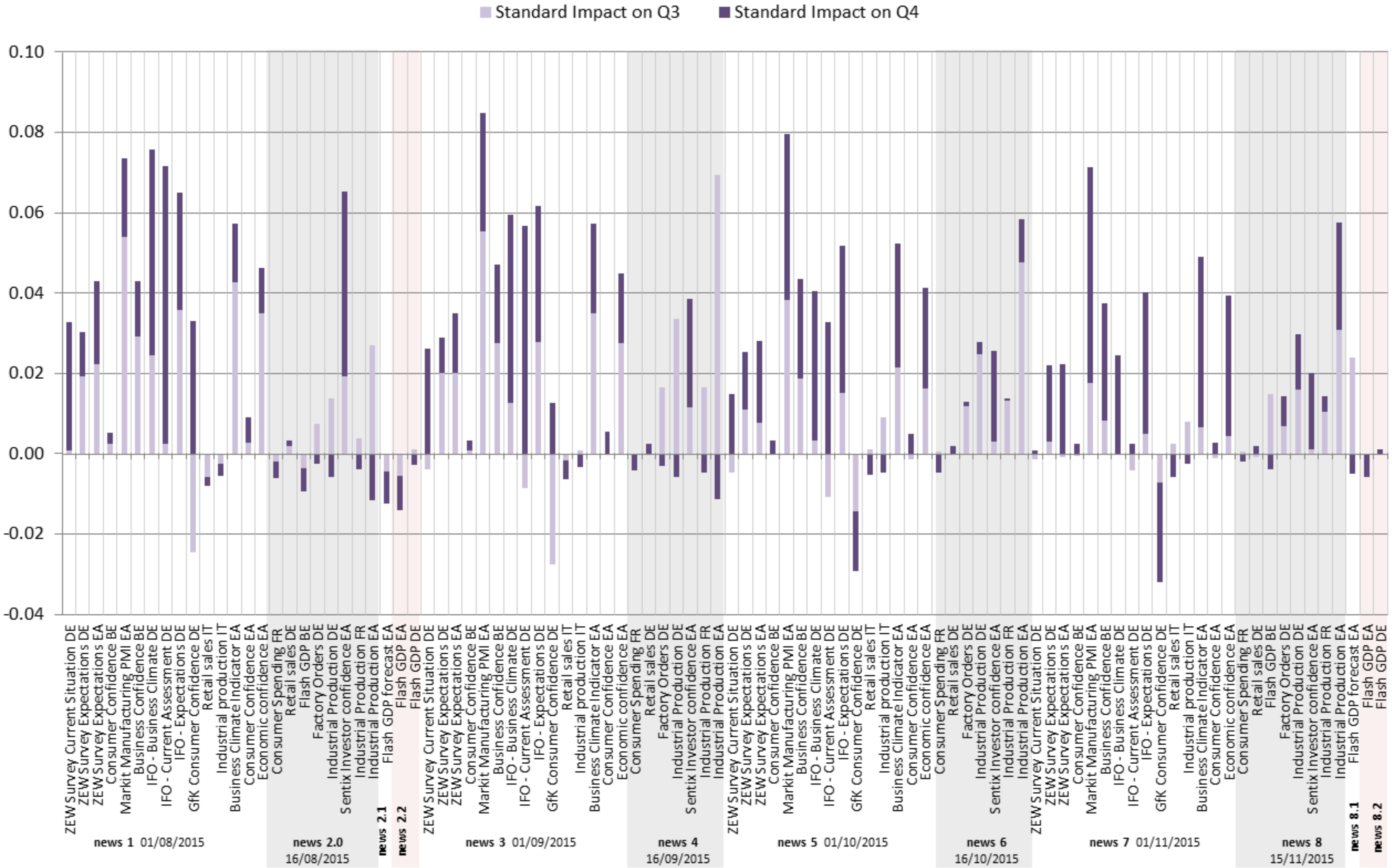
The graph can be read in a more detailed way. For example, inside the first block of news, we find the Markit PMI release that is *read* by our model on the first of August (but the information it holds still refers to the month of July). This publication, which takes place in the second month of the current quarter, is expected to induce an upward or downward revision of real GDP growth expectations in the euro area close to 0.06 for that quarter, not surprisingly, but also 0.02 for the next quarter already. Markit's euro area PMI release could thus clearly be considered as a timely indicator. The graph also reveals that GDP growth expectations for Q3 will remain largely affected by the August, September and October releases of the euro area PMI as well, although the impact is gradually decreasing. Interestingly, Figure 6 also demonstrates that the industrial production release in Germany and the euro area have a non-trivial standard impact even if they are published with a delay larger than one month. This result may question the empirical findings of Camacho and Pérez-Quirós (2010) or de Antonio Liedo (2015), where the impact of hard data is found to be very small in favour of surveys. Our results rely on our modelling approach based on a relatively large number of factors, which aim to give the model a fair chance to fit a very heterogeneous dataset that combines surveys coming from multiple sources together with the hard data expressed in terms of both monthly and yearly growth rates. The graph illustrates that the industrial production release in the block of news 4, which corresponds to the first month of Q3, i.e. July, still has a very large impact on the euro area GDP for that quarter. The standard impact of 0.07 for industrial production in the euro area turns out to be comparable to the sum of the impacts of the other hard indicators combined.

This graph also reveals that the Belgian flash GDP release (within the block of news 8.0) has a very small impact on aggregate euro area growth referring to the same quarter, even though it is published two weeks earlier. The estimated impact on euro area expectations is very small, which suggests that it does not incorporate much added value beyond all indicators that have been previously released.

Despite the irrelevance of the Belgian flash GDP publication, the results do prove that certain national indicators can nonetheless be useful for the short-term prediction of the euro area aggregate business activity. The Business Confidence indicator published by the National Bank of Belgium, for example, turns out to be among the releases with the largest impact, together with the IFO surveys for Germany.

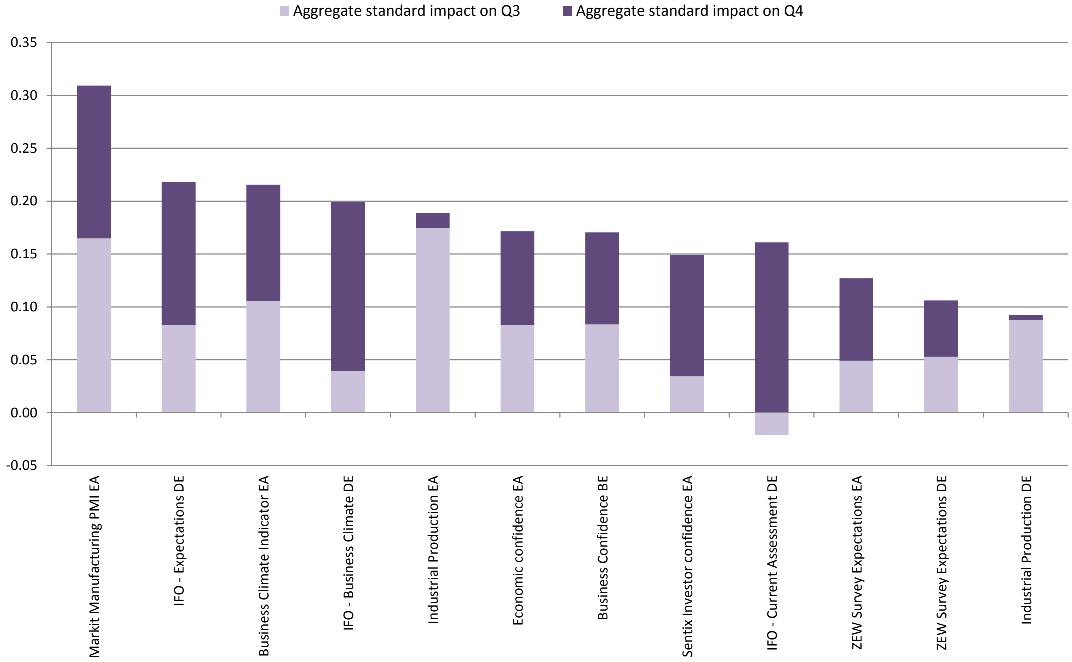
A ranking of the releases according to their cumulative impact over a whole semester is represented in Figure 7. The figure confirms the dominance of soft data, but industrial production in the euro area and Germany occupy the fifth and twelfth place, respectively, in the overall ranking. Two remarks are necessary to interpret our results. First, by analysing the data releases by blocks, we are neglecting the fact that Markit euro area PMI and the NBB Business Survey are published somewhat earlier than the competing indicators. Had we taken that aspect into account, the standard impact associated to both indicators would have been even larger. Second, this ranking is constructed using standard impacts that depend on how much each piece of news is weighted by the model for the prediction of our target variable, euro area GDP (see Figure 3). The weights associated to the news are a function of the model parameters, and more specifically, of their correlation with the target variable. This implies that if the target variable would be German GDP instead of euro area GDP, the resulting ranking could be different (see Section 4).

Figure 6: Standard Impacts for euro area GDP Flash



This figure represents the weigths associated to the real-time newsflow multiplied by the standard deviation of each news.

Figure 7: Ranking According to the Standard Impacts for euro area GDP Flash



This graph provides some sort of a 'horizontal summation' of the impacts that occur in Figure 6, i.e. for every indicator, the sum is made of the impact of the news that is being read between the first of August and mid-November. Only the twelve highest-ranked indicators are shown. The distribution of the colors in each bar (light vs dark) also gives a first indication about the timeliness of the indicator. The fact that news about the hard data mainly impacts the expectation for Q3 is a reflection of the release calendar (Figure 2), as hard data that can be read by the model between August and mid-November actually still refer to the months between May and September.

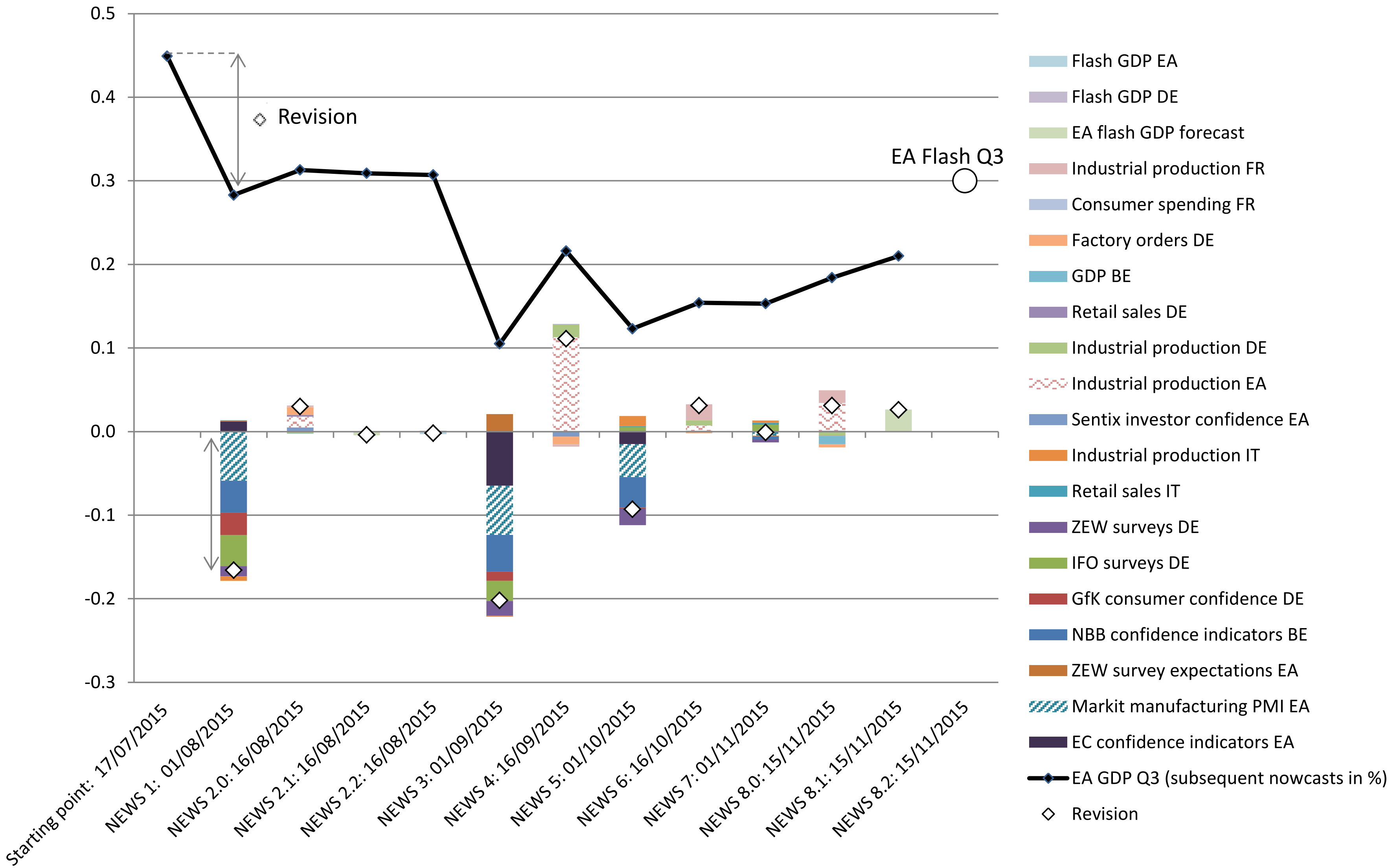
3.5 Real-Time Monitoring and Evaluation

The standard impacts of the news that we have represented in Figure 6 provided us with a measure of the average contribution of the news to the nowcasts updating the growth expectations. In practice, the news is unpredictable and will follow a normal distribution with a given standard deviation. In our example, the process of updating GDP growth expectations for 2015Q3 in the euro area can be decomposed in terms of the realized news. Thus, Figure 8 represents the process of updating GDP on the basis of the model's automatic interpretation of the real-time dataflow.

Once we have understood how the model works in practice, it is worth investigating how reliable the model forecasts are:

- Over a long sample, we expect the forecasting accuracy of the model to improve after reading each block of news. Which news blocks are the most significant updates? Please refer to Table A.2 in the Annex. Even though every update brings a decrease in RMSE, only some of them prove to be statistically significant. Table A.2 also shows the results of two encompassing tests. In the first case, the null hypothesis

Figure 8: Real-Time Updates for 2015Q3



This figure represents the updating process of the GDP nowcast every time news arrives. In order to facilitate the reading of the figure, certain variables were taken together (e.g. IFO surveys DE is compiled of three separate IFO indicators).

states that the updated forecast encompasses all the relevant information from the old forecast. In the right column of the Table, the null hypothesis works just the other way around. As expected, in most of the cases, one could reject that the old forecast encompasses the updated one. This confirms that the updated forecast adds significant information that is not present in the previous forecast. When comparing our DFM nowcast at the +44 days-horizon (i.e. right before the flash release) to the one published by Bloomberg, the latter yields a significantly better RMSE so we have to reject the null of equal forecasting accuracy. However, interestingly, neither of them seem to encompass the other, which means that they both contain complementary information and the forecasts could be improved by combining these two models together.

- Our nowcasts at different points in time (middle of the quarter, end of the quarter and one day prior to the flash release) can be compared with those coming from Now-Casting.com, an economic forecasting business that publishes their nowcasts online. Other important benchmarks are the PMI-implied growth rates for a given quarter[10] or the Bloomberg forecast that is available one day before the flash release. A Table of comparison (A.3) can be found in the Annex. Our model appears to register significantly worse forecasts than the Now-Casting.com benchmark at the middle of the quarter (-45 days), but this could be considered to be rather normal as their platform makes use of a much larger dataset. One day prior to the flash release (+44 days), our model significantly improves over the Now-Casting.com benchmark, but according to the encompassing test, they both contain valuable information that should not be neglected. The second part of the table suggests that Bloomberg and the PMI rule are comparable to our model in terms of accuracy. The first encompassing test suggests that both benchmarks contain useful information that is not present in our DFM nowcast. But at the same time, the second encompassing test indicates that the DFM contains useful information that is not captured by the PMI-implied forecast, nor by the Bloomberg benchmark.

[10]Please refer to the publication by Markit Economic Research entitled 'Using PMI survey data to predict official eurozone GDP growth rates' to find out more about these PMI-implied growth rates.

4 Robustness

In this section, we investigate the robustness of our benchmark results (referred to as Analysis A) regarding the ranking when the target variable is German GDP flash instead of the euro area flash release (Analysis B). We will also study whether the large impact we have found for surveys remains under the assumption that hard data are published without delay (Analysis C). Furthermore, we will re-calculate the results under the hypothesis that hard data is published without revisions. That is, instead of the original data releases, which contain preliminary data that will be subject to significant revisions in the case of hard data, we use the series that have already been revised (Analysis D). Finally, we investigate the *combined* impact of having fully revised hard data that are published without any delay on fully revised GDP (analysis E).

Table 2: Robustness Analysis

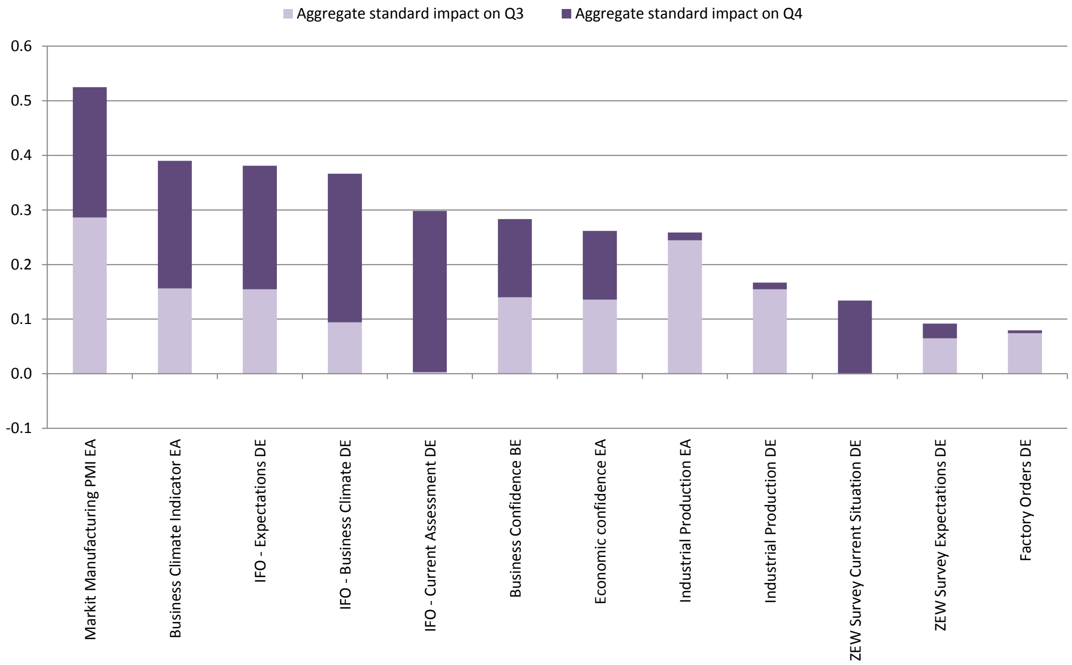
Analysis	Objective	Details	Results
A	Impacts on euro area GDP flash (see subsection 3.4)	Real-time dataflow	Figures 6 and 7
B	Impacts on German GDP flash	Real-time dataflow	Figure 9
C	Counterfactual impact on euro area GDP flash	Hard data published without any delay	Figure 10
D	Counterfactual impact of revised hard data on revised euro area GDP	Hard data are revised, but published according to the actual real-time calendar	Figure 11
E	Counterfactual impact of revised hard data on revised euro area GDP	Hard data is fully revised and published without any delay	Figure 12

4.1 Standard impacts when the target becomes German GDP

In this section, we re-calculate the standard impacts depicted in Figure 6 and the resulting ranking in Figure 7 in the case that our target is German flash GDP instead of the euro area flash. The ranking of indicators is shown in Figure 9. The top four of best-ranked indicators remains unchanged, lead by the Markit PMI. Industrial production in the euro area loses a few spots in the ranking, but remains in the top ten. Industrial production in Germany is now following more closely that of the euro area, in terms of ranking. The

NBB Business Confidence has moved to the sixth position after the IFO Business Climate and Expectations for Germany.

Figure 9: Ranking According to the Standard Impacts for German GDP

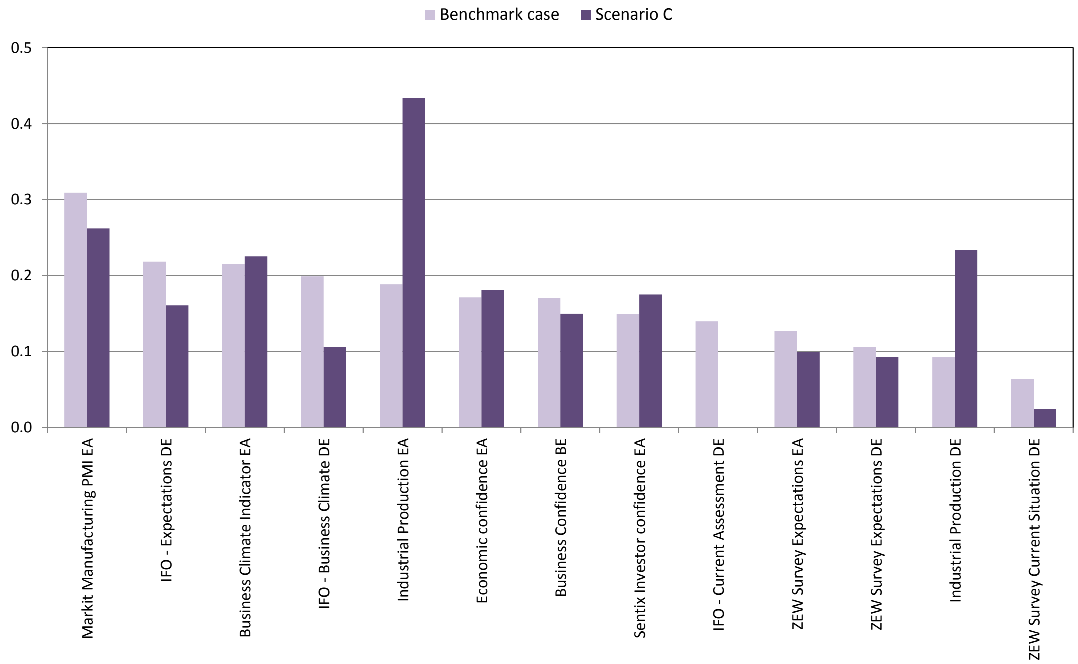


This figure shows the results of robustness analysis B (cf. scenarios described in Table 2). Only the twelve highest-ranked indicators are shown.

4.2 Standard impacts when we incorporate the timeliness advantage in all hard indicators

Timeliness is a unique characteristic of soft data that probably contributes to their large impact, as discussed in the benchmark case. In this counterfactual exercise, we analyse if (and by how much) those impacts are reduced when we assume that hard data are published already at the end of the reference month, i.e. the publication delay equals zero. Figure 10 resumes the benchmark ranking following the usual format and compares this to the impact that emerges from the counterfactual exercise. The first thing that strikes the eye is that one of the variables, industrial production in the euro area, has now obtained a very large impact. Industrial production in Germany also gains some importance, and its standard impact is now close to that of the manufacturing PMI indicator for the euro area. Industrial production in Italy or France have smaller impacts, in spite of their counterfactual timeliness. Surprisingly, the sum of the impacts of all soft indicators in Figure 10 is still larger than the aggregated impact of all hard indicators that are represented in the ranking. This implies that timeliness is not the only characteristic of soft data contributing to their large impact. This result has also been discussed by de Antonio Liedo (2015) in the context of Belgian data using a different model, although the role of hard indicators is practically negligible according to his findings. Our result does confirm the conclusion by Piette and Langenus (2014), who find that even when all hard indicators are available, survey data still contain relevant information that is not captured by the usual set of hard data. Gayer et al. (2015) argue more specifically that survey data have other characteristics besides their timeliness that can possibly improve the nowcasting accuracy: they are often forward-looking and also tend to have a broader sectoral scope.

Figure 10: Ranking According to the Counterfactual (timeliness) Standard Impacts for euro area GDP



This figure shows the results of the benchmark case (light purple) and those of robustness analysis C (dark purple). Note that each bar represents the aggregated standard impact over an entire semester (i.e. the impact on both Q3 and Q4).

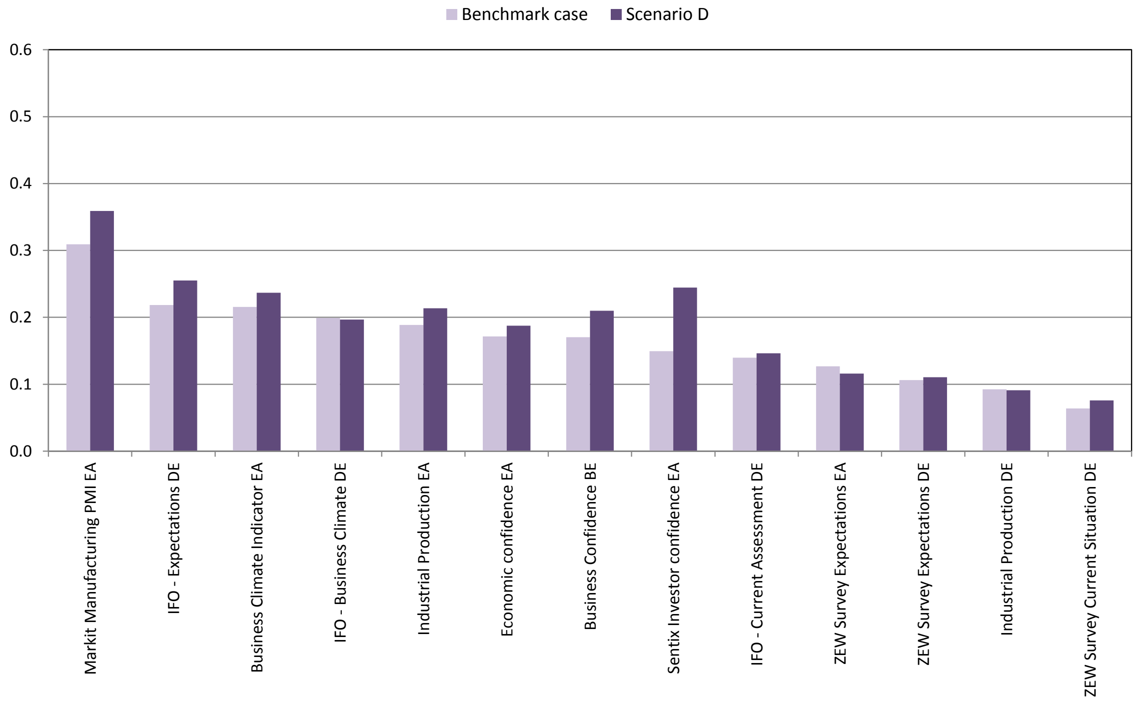
4.3 Standard impacts on revised GDP when we incorporate revisions in all hard indicators

An important innovation of this paper is the fact that it makes use of the real-time dataflow, directly obtained from press statements. To the best of our knowledge, earlier work on nowcasting usually deals with a *pseudo real-time* dataset that makes use of the most recent (i.e. already revised) data, that are treated as if they were real-time (i.e. the dataset also has a ragged-edge, because the publication delay is the same as in the benchmark case). Robustness analysis D sort of mimics such a pseudo real-time dataset: the growth rates for the hard indicators were replaced by the most recent data found via Thomson Reuters Datastream and are therefore considered *revised*. It may be important to note that only the month-on-month growth rates were replaced by their revised counterpart, while the year-on-year growth rates were kept unchanged (i.e. first vintage). Otherwise, this would imply feeding twice the same information into the model and this could cause the dynamic factor model to attribute an abnormally large weight to certain hard indicators.[11]

Figure 11 resumes the ranking that emerged from the benchmark case, and combines it with the standard impacts of the current counterfactual analysis D. The comparison between these two series may prove to be somewhat tricky, because the impacts for the benchmark case were calculated with regard to euro area flash GDP, while the impacts for this counterfactual analysis D were calculated with regard to euro area revised GDP. After all, researchers who have constructed a pseudo-real time set and are using the most recent hard data, will most likely also use the most recent (i.e. revised) GDP variable, and it is therefore the only relevant target variable to use in this counterfactual exercise. Our exercise proves that analyzing the release impacts based on a pseudo real-time dataset may be misleading. The impacts are not informative about the actual influence of news on the way agents form their expectations, for two reasons: first, the news component of the data releases distilled in real-time does not fully coincide with the forecast errors based on the pseudo real-time dataset, and second, the relevant influence of news on the genuine real-time target, the flash GDP, may be biased if the revised series of GDP are used instead.

[11]This was less of a risk in the benchmark case, because the annual growth rate refers to a level of industrial production of one year ago, which is likely to already be revised. Hence, while the monthly growth rate only provides information on the most recent observation, the annual growth rate already gives an indication of past revisions.

Figure 11: Ranking According to the Counterfactual (revisions) Standard Impacts for euro area revised GDP

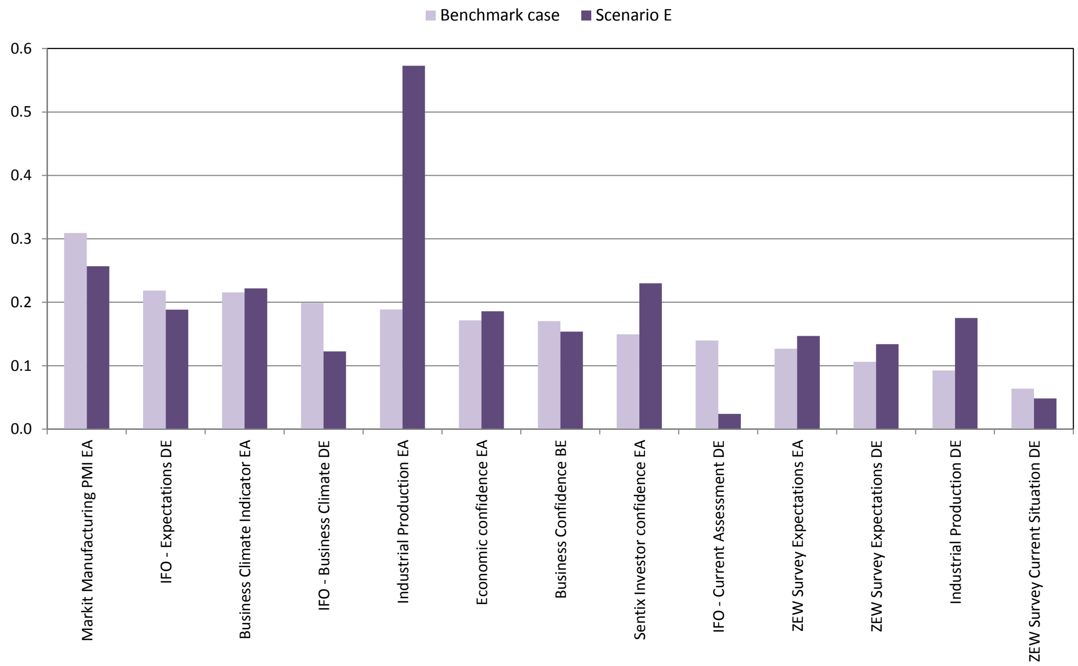


This figure shows the results of robustness analysis D (cf. scenarios described in Table 2), compared to the benchmark results. Note that each bar represents the aggregated standard impact over an entire semester (i.e. the impact on both Q3 and Q4).

4.4 Standard Impacts on revised GDP when we incorporate the timeliness advantage and revisions in all hard indicators

The dataset in the current robustness analysis consists of the most extreme scenario: all hard data are revised, while they are also assumed to be published without any delay. The resulting ranking is based on the standard impact of the news on the model's estimate for *revised* euro area GDP. Figure 12 confirms that industrial production in the euro area is the big winner in this scenario. What is more surprising, is the observation that the survey indicators also maintain their significant impact for the prediction of revised GDP. This is an important result, because it may appear to be somewhat contradictory with the way that "final" (i.e. revised) GDP is supposed to be assembled. According to the producers of the national accounts, only hard data are taken into account when the revised GDP series are being made. Hence, survey data should prove to be irrelevant for the revised euro area GDP series. We see two possible reasons to explain why this is not the case. First, for some reason, statisticians may prefer to keep their final GDP estimate as close as possible to the flash GDP estimate. As the flash estimate relied mainly on survey variables, the impact of surveys may simply be propagated onto the final GDP as well. Second, when hard data are behaving extraordinarily, and statisticians suspect that something is wrong with the input series, they may decide to apply some *judgment* to the mechanical estimates. In this case, our results would show that statisticians' expert judgment is influenced by the information from soft data.

Figure 12: Ranking according to the Counterfactual (timeliness + revisions) Standard Impacts for *revised* euro area GDP



This figure shows the results of robustness analysis E (cf. scenarios described in Table 2), compared to the benchmark results. Note that each bar represents the aggregated standard impact over an entire semester (i.e. the impact on both Q3 and Q4).

4.5 Variable selection based on expected impacts

The reader of this paper may be tempted to use this methodology in order to identify a reduced set of variables for the estimation of a small dynamic factor model. This idea could be considered as a refinement of the method proposed by Rünstler (2016). Rünstler's analysis is based on earlier work by Bańbura and Rünstler (2011) and exploits the weights of the different predictor variables[12] in the factors.

In their set-up, those weights are given by a measure of the historical correlations of the *revised* predictor variables with the factors extracted from them. This means that the

[12]See Harvey and Koopman (2003) for details on the calculation of observation weights.

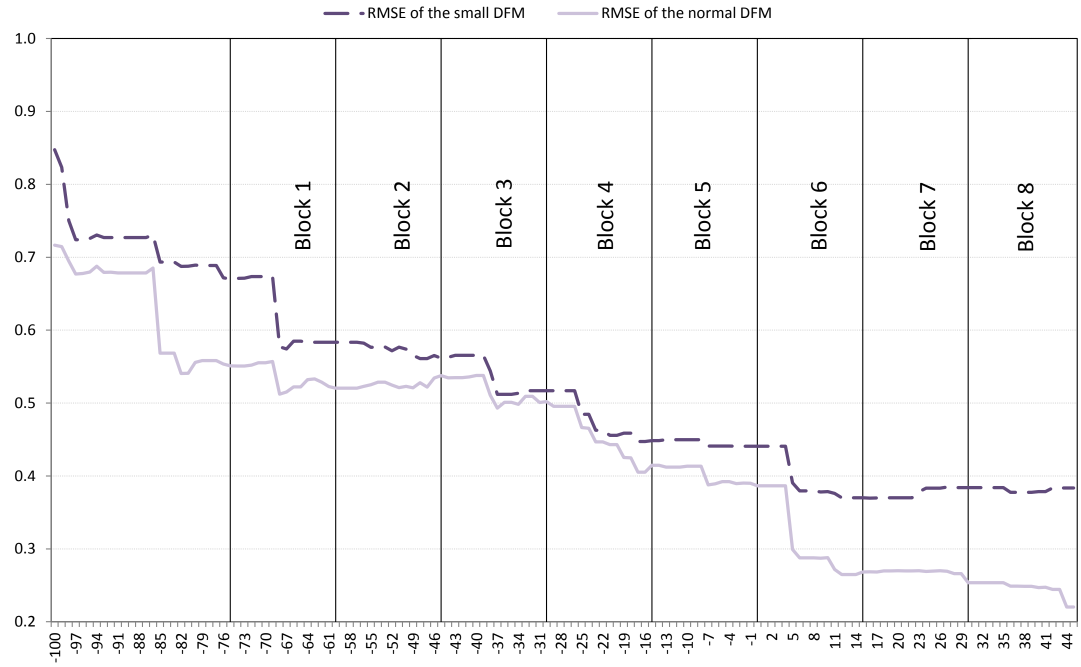
results may be distorted by data revisions that have taken place years after the actual data releases. Second, from a theoretical point of view, selecting variables in function of those weights can also be misleading when the correlation across predictor variables is neglected, as it is the case in Rünstler's econometric framework. This weakness is visible in the analysis by Camacho and Pérez-Quirós (2010), who exploit the same idea. These authors have noted, in the context of their 'Eurosting' model, that the euro area GDP forecast obtained on January the 24th of 2007 for the first quarter of that year was totally driven by the NBB Business Survey, simply because it was the only indicator available for January. This can be misleading because that figure does not necessarily change the forecast and it could be anticipated to some extent on the basis of other indicators that were available.

Regardless of the underlying methodology, the idea is to select the indicators with the highest weight in the forecasts. In our set-up, we select the indicators that induce the largest updates in the model's forecasts. Thus, we ensure that timeliness does not bias our analysis of contributions to the detriment of quality by considering the correlation patterns across all news as an essential part of our modelling framework. A second difference with respect to Rünstler's approach is that we perform the analysis in a genuine real-time environment, whereas he operates within a pseudo-real time dataset.

Although the idea of variable selection is appealing, and it has worked in Rünstler's laboratory, it has not worked in our case. As shown in Figure 13, selecting the highest-ranked indicators deteriorates the forecasting performance of the model. The RMSE function corresponding to our workhorse model with four factors deteriorates remarkably when the same model is re-estimated using only the highest-ranked variables. In particular, the reduction in the RMSE that takes place as a consequence of the news defined in block 6, is now smaller, and subsequent updates do not improve the forecasts.

The phenomenon reported above can be easily explained. Our dynamic factor model was originally estimated via maximum likelihood using the full set of variables. Thus, the factors are determined in such a way that they help to match the dynamics of all series, including those that turn out to have a smaller weight for GDP, which we have discarded along the lines of Rünstler (2016). However, those series that we have disregarded may be crucial to forecast the series that actually have a large impact for GDP. Given the complex interactions between all series, which we aim to capture with a 4 factors model, using the standard impacts on GDP as a criterion to discard the series may not be a good idea.

Figure 13: RMSE functions of the normal model and the small model



5 Conclusion

This paper provides a formal way to quantify the incremental value of alternative business cycle indicators that are often monitored for nowcasting GDP growth in the euro area. The objective is to rank all those indicators according to their importance for the growth forecasts in order to facilitate the complex task of interpreting such an heterogeneous and asynchronous flow of data releases. To do so, we propose a state-space representation for a dynamic factor model with a sufficiently large number of factors in order to account for the joint dynamics of all the series, which are constructed using the first available data vintage as reflected in the original press releases. Such a representation allows us to define the news or unexpected component of each data release in terms of the Kalman filter innovations. Their precise impact on the GDP growth estimate is determined by both the timeliness and quality of the news, which is captured by the Kalman gain. Thus, GDP forecasts are updated after interpreting the news or surprise component of each data release. The model-based news is conceptually equivalent to the forecast errors made by analysts monitoring the data releases.

Gayer et al. (2015) also exploit the framework of dynamic factor models and introduce a stylized control for the timeliness of the various blocks of indicators, which is standard practice since Giannone et al. (2008), but they use a totally different methodology to study the role of surveys. Hence, we want to stress the various aspects of our methodology that represent a novel and powerful approach to think about the real-time impact of predictor variables on a given target. First, we believe that time series econometrics that rely of the revised history of a given indicator cannot answer the same question due to the large size of data revisions in certain series such as industrial production or sales. Our empirical results are based on time series constructed from real-time press releases in order to correct for artificial correlation patterns that may be present in historical time series in both surveys and hard data. This implies that in contrast to standard real-time evaluations based on the historical time-series available at a given point in time, our results will not be artificially distorted by revisions, such as seasonal adjustments, redefinitions or reweighing, that are sometimes executed with the benefit of hindsight.

Second, we state that the specific formulation of our research question in this paper (determining the *incremental* information content of a given press release, rather than of a block of releases as is the case in, for example, the paper by Gayer et al. (2015)) is a rather unique approach. While earlier literature has indeed already confirmed the importance of a given block of certain indicators (e.g. surveys) in order to reduce forecast errors, this information does not provide us with any clue regarding the impact of each one of the elements in the block. After all, a few particular surveys may be determining the performance of the whole block.

More specifically, we find that, in the process of updating nowcasts for euro area GDP growth, the strongest impact corresponds to the Markit Manufacturing PMI and the Business Climate Indicator for the euro area, followed by the IFO Business Climate and IFO Expectations for Germany. Interestingly, the NBB's own business confidence indicator for Belgium is following closely those survey variables and obtains the seventh place in the overall ranking. More generally, it is quite remarkable that none of the consumer confidence indicators occur in the overall ranking. These findings have not been presented before. When it comes to hard data, euro area industrial production occupies the fifth place in that ranking and is actually the only hard variable that makes it into the top ten. In the counterfactual scenario where hard data for a given month would be released exactly at the end of that month, industrial production in the euro area and Germany would rank first and third, respectively, while the Manufacturing PMI for the euro area would still have the second largest impact. This suggests that, in addition to being available in a more timely manner, survey data also contain relevant information that is not captured by hard data. Having an overview of the impact of the different data releases helps to make the (use of) models more transparent for the user, but as shown in section 4.5, it is not necessarily a good idea to simply rely on our ranking as a data selection tool.

References

[1] Abberger, K. (2007). "Qualitative business surveys and the assessment of employment – a case study for Germany". *International Journal of Forecasting*, **23**, 249-258.

[2] Angelini, E., G. Camba-Mendez, D. Giannone, G. Rünstler and L. Reichlin (2011). "Short-Term Forecasts of Euro Area GDP Growth". *The Econometrics Journal*, **1**, C25-C44.

[3] Aruoba, S.B., Diebold, F.X. and C. Scotti, 2009. "Real-Time Measurement of Business Conditions". *Journal of Business and Economic Statistics*, American Statistical Association, **27(4)**, 417-427.

[4] Bańbura M. and M. Modugno (2010). "Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data". *ECB Working Paper Series*, **1189**

[5] Bańbura M., D. Giannone and L. Reichlin (2011). "Nowcasting". *Oxford Handbook on economic Forecasting*, in Michael P. Clements and David F. Hendy.

[6] Bańbura M., D. Giannone, M. Modugno and L. Reichlin (2013). "Nowcasting and the Real-Time Data Flow". *Handbook of Economic Forecasting*, in G. Elliot and A. Timmermann, **2**, Elsevier.

[7] Camacho M. and G. Pérez-Quirós (2010). "Introducing the Euro-Sting". *Journal of Applied Econometrics*, **25**, 663-694.

[8] Claveria, O., Pons, E. and R. Ramos (2007). "Business and consumer expectations and macroeconomic forecasts". *International Journal of Forecasting*, **23**, 47-69.

[9] D'Agostino, A. and B. Schnatz (2012). "Survey-based nowcasting of US growth: a real-time forecast comparison over more than 40 years". *ECB Working Paper Series*, **1455**.

[10] de Antonio-Liedo (2015). "Nowcasting Belgium". *Eurostat review on National Accounts and Macroeconomic Indicators* , **75**, 7-48

[11] Dempster, A.P., N.M. Laird and D.B. Rubin (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society. Series B (Methodological)*, **39:1**, 1-38.

[12] Diebold F.X. and R.S. Mariano (1995). "Comparing Predictive Accuracy". *Journal of Business and Economic Statistics*, **13**, 253-265.

[13] Gayer, A., A. Girardi and A. Reuter (2015). "The role of survey data in nowcasting euro area GDP growth". *International Journal of Forecasting*, **35**, 400-418.

[14] Koenig E. F. (2002). 'Using the Purchasing Managers' Index to Assess the Economy's Strength and the Likely Direction of Monetary Policy". *Federal Reserve Bank of Dallas Economic and Financial Policy Review*, **1**, 2-14.

[15] Lui, S., Mitchell, J. and M. Weale (2011)."Qualitative business surveys: signal or noise?". *Journal of the Royal Statistical Society: Series A*, **174**, 327-348.

[16] Doz C., D. Giannone and L. Reichlin (2012). "A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models". *Review of Economics and Statistics*, **94**, 1014-1024.

[17] Durbin, J and S.J. Koopman (2001). "Time Series Analysis by State Space Methods ". Oxford University Press.

[18] Giannone, D., L. Reichlin and D. Small (2008). "Nowcasting: The Real-Time Informational Content of Macroeconomic Data Releases". *Journal of Monetary Economics*, **55**, 665-676.

[19] Giannone, D., L. Reichlin and S. Simonelli (2009). "Nowcasting Euro Area Economic Activity in Real Time: The Role of Confidence Indicators". *National Institute Economic Review*, **210**, 90-97.

[20] Kalman, R.E. (1960). "A new approach to linear filtering and prediction problems". *Journal of Basic Engineering*, **82**, 35-45.

[21] Kishor, N.K. and E.F. Koenig (2012). "VAR Estimation and Forecasting When Data are Subject to Revision". *Journal of Business and Economic Statistics*, **30**, 182-190.

[22] Martinsen K., F. Ravazzolo and F. Wulfsberg (2014). "Forecasting Macroeconomic Variables Using Disaggregate Survey Data". *International Journal of Forecasting*, **30**, 65-77.

[23] Piette C. and G. Langenus (2014). "Using BREL to now-cast the Belgian business cycle: the role of survey data". *Economic Review*, National Bank of Belgium, 75-98.

[24] Mariano R.S. and Y. Murasawa (2003). "A new coincident index of business cycles based on monthly and quarterly series". *Journal of Applied Econometrics*, **18**, 427-443.

[25] Rünstler, G. (2016). "On the Design of Datasets for Forecasting with Dynamic Factor Models". *ECB Working Paper Series*, **1893**

[26] Shumway, R. and D. Stoffer (1982). "An approach to time series smoothing and forecasting using the EM algorithm". *Journal of Time Series Analysis*, **3**, 253-264.

Annex - Evaluating Forecasting Accuracy

The prediction errors are defined with a reference i to the information set available at the time the forecast was made:

$$e_{t|i} = y_t - \hat{y}_{t|\mathcal{F}_i} \tag{A.1}$$

where $\mathcal{F}_i$ need not only include lags of y_t. In practice, the information that will be actually used may be a small subset of $\mathcal{F}_i$.

The properties of these forecast errors can be assessed in isolation or relative to a benchmark, which we will define as $\breve{e}_{t|i}$. The benchmark may be a naive forecast, e.g. random walk, in which case $\breve{y}_{t|\mathcal{F}_i}$ would be equal to $\breve{y}_{t|y_{t-1}} = y_{t-1}$. However, the benchmark could also be a prediction regularly published by a forecasting institute or market analysts, i.e. Bloomberg, which is not necessarily model-based. In that case, $\breve{y}_{t|\mathcal{F}_i}$ would be given by methods and a subset of $\mathcal{F}_i$ which is unknown to us.

For model-based forecasts, we use the following notation: $\hat{y}_{t|\mathcal{F}_i} = E_\theta[y_t|\mathcal{F}_i]$ to highlight the fact that they are based on model-consistent expectations given by the parameter vector θ.

In forecasting comparisons involving competing forecasts resulting from the same information set, the subindex i will be removed because it does not play a role. We will first test the following hypothesis involving forecast errors:

$$Unbiasedness: \quad E[e_t] = 0 \tag{A.2}$$

$$Autocorrelation: \quad E[e_t e_{t-1}] = 0 \tag{A.3}$$

$$Equality\ in\ squared\ errors: \quad E[e_t^2 - \breve{e}_t^2] = 0 \tag{A.4}$$

$$Equality\ in\ absolute\ errors: \quad E[|e_t| - |\breve{e}_t|] = 0 \tag{A.5}$$

$$Forecast\ \hat{y}_t\ encompases\ \breve{y}_t: \quad E[(e_t - \breve{e}_t)e_t] = 0 \tag{A.6}$$

$$Forecast\ \breve{y}_t\ encompases\ \hat{y}_t: \quad E[(\breve{e}_t - e_t)\breve{e}_t] = 0 \tag{A.7}$$

Diebold-Mariano Test

The test originally proposed by Diebold and Mariano (1995) considers a sample path of loss differentials $\{d_t\}_{t=1}^T$. In the case of a squared loss function, we have $d_t = e_t^2 - \breve{e}_t^2$.

Table A.1: Forecasting Evaluation Tests

Test	Null hypothesis	Statistic	Asym. theory	Finite sample
Bias	$E[e_t] = 0$	$\text{B} = \dfrac{\bar{e}}{\sqrt{\dfrac{2\pi \hat{f}_e(0)}{T}}}$	$N(0,1)$	KV(2005)
Autocorrelation	$E[e_t e_{t-1}] = 0$	$\text{AR} = \dfrac{\bar{\rho}}{\sqrt{\dfrac{2\pi \hat{f}_\rho(0)}{T}}}$	$N(0,1)$	KV(2005)
Diebold-Mariano	$d_t \equiv L_{1,t} - L_{2,t} = 0$	$\text{DM} = \dfrac{\bar{d}}{\sqrt{\dfrac{2\pi \hat{f}_d(0)}{T}}}$	$N(0,1)$	KV(2005)
Encompassing 1	$d^e_{1,t} \equiv E[(e_t - \breve{e}_t)e_t] = 0$	$\text{E}_1 = \dfrac{\bar{d}_1}{\sqrt{\dfrac{2\pi \hat{f}_{d_1}(0)}{T}}}$	$N(0,1)$	KV(2005)
Encompassing 2	$d^e_{2,t} \equiv E[(\breve{e}_t - e_t)\breve{e}_t] = 0$	$\text{E}_2 = \dfrac{\bar{d}_2}{\sqrt{\dfrac{2\pi \hat{f}_{d_2}(0)}{T}}}$	$N(0,1)$	KV(2005)

Under the assumption that the loss differential is a covariance stationary series, the sample average, $\bar{d}$, converges asymptotically to a normal distribution:

$$\sqrt{T}\bar{d} \quad \underset{\rightarrow}{d} \quad N(\mu, 2\pi f_d(0)) \tag{A.8}$$

In particular, they proposed to test the null hypothesis that the forecast errors coming from the two forecasts bring about the same loss: $E[e_t^2 - \breve{e}_t^2] = 0$ against the two-sided alternative. Thus, the resulting p-values represent the probability of obtaining the realized forecast error differential or a more extreme one in a new experiment if the null hypothesis was actually true. The test-statistic that will be used to calculate our p-values is computed as follows:

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{T}}} \tag{A.9}$$

where $2\pi \hat{f}_d(0)$ is a consistent estimate. Consider $2\pi \hat{f}_d(0) = \sum_{\tau=-(T-1)}^{(T-1)} w_\tau \gamma_d(\tau)$, where $\gamma_d(\tau) = \frac{1}{T}\sum_{t=|\tau|+1}^{T}(d_t - \bar{d})(d_{t-|\tau|} - \bar{d})$. Under the assumption that $gamma_d(\tau) = 0$ for $\tau \geq h$, we can use a rectangular lag window estimator by setting $w_\tau = 0$ for $\tau \geq h$. Another option is to use the Heteroschedasticity and Autocorrelation Consistent (HAC) estimator proposed by Newey and West (1987). In this case, the weights could be given by a triangular window, $w_\tau = 1 - \frac{\tau}{h}$ for $\tau < h$. In this case, however, the consistency property only remains valid when the truncation lag h or bandwidth is a function of the sample size T.

The idea is to test the statistical significance of the regression of $e_t^2 - \breve{e}_t^2$ on an intercept. In order to determine the statistical significance of the intercept, its associated standard errors need to take into account the autocorrelation patterns of the regression error, which are considered in the denominator of equation (A.9). JDemetra+ exploits the same unified framework to conduct all tests listed in Table A.1. But given the small sample sizes that are typical in real-time forecasting applications, which leads to an over-rejection of the null hypothesis, we follow Coroneo and Iacone (2015) and use a finite sample distributions of Kiefer and Vogelsang (2005). The distribution of the test statistic (A.9) will depend on kernel and the bandwidth chosen, which is set by default equal to $T^{0.5}$. The results can be very different than those resulting from the traditional asymptotic theory, where the test statistic would have the same distribution under the null independently of the kernel and the bandwidth used.

Table A.3 contains the results of this test together with the encompassing test and two efficiency tests, which are described below.

Encompassing Test

Independently of whether the null hypothesis $E[e_t^2 - \breve{e}_t^2] = 0$ is rejected or not, it is relevant to understand to what extent our model encompasses all the relevant information of the benchmark, and the other way around. Because of the obvious symmetry of both statements, we consider only the first one. If our forecasts $y_{t|\mathcal{F}_i}$ encompass a given benchmark $\breve{y}_{t|\mathcal{F}_i}$, the difference between those benchmark forecasts and ours will not be a relevant factor in explaining our own forecast error. In other words, the regression coefficient λ will not be significantly different from zero in the following regression:

$$\underbrace{y_t - y_{t|\mathcal{F}_i}}_{e_t} \quad = \quad \lambda \underbrace{(\breve{y}_{t|\mathcal{F}_i} - y_{t|\mathcal{F}_i})}_{e_t - \breve{e}_t} + \xi_t \tag{A.10}$$

$$\Updownarrow$$

$$y_t = \lambda \breve{y}_{t|\mathcal{F}_i} + (1-\lambda) y_{t|\mathcal{F}_i} + \xi_t \tag{A.11}$$

Following Harvey, Leybourne and Newbold (1998), the statistical significance of the λ coefficient in regression A.10 can be used to reject the null hypothesis that our model encompasses the benchmark. In this case of rejection, equation A.11 suggests that a combination of the two forecast would yield a more informative forecast.

By construction, the value of the coefficient of a regression $\breve{e}_t = \alpha \breve{e}_t - e_t + \xi_t$ is equal to $1 - \lambda$, but it is not necessarily true that the *rejection* of the null hypothesis in the first case implies the *acceptance* of the symmetric statement.

The test-statistic is computed as follows. When the null hypothesis is that our model encompasses the benchmark, we define the sequence $\{d_t\}_{t=1}^T$, where $d_t = e_t(e_t - \breve{e}_t)$, and we compute $E1 = \dfrac{\bar{d}}{\sqrt{\dfrac{2\pi \hat{f}_d(0)}{T}}}$, exactly as in equation A.9.

Efficiency: Bias Test

In order to assess whether our forecasts are unbiased, we will simply test the statistical significance of the average error. In some cases, the time series of forecast errors $\{e_t\}_{t=1}^{T}$ may be autocorrelated to some extent even when they are based on a model with IID innovations. In such cases, the variance associated to the estimate of the average forecast error may be large. The test statistic has exactly the same form as the previous tests discussed so far.

Efficiency: Autocorrelation Test

We will test here a second necessary condition for our forecasts to be efficient: absence of autocorrelation. In the same spirit as the tests described above, we will assess the statistical significance of the forecast errors' autocorrelation. Thus, our sequence $\{d_t\}_{t=1}^{T}$ will be defined with $d_t = e_t e_{t-1}$.

Testing the rationality of nowcasting updates

Patton and Timmerman (2012) suggest testing whether the mean squared forecast error is actually decreasing when the horizon decreases. This idea could be applied in our set-up by replacing the concept of forecast horizon with the number of days from the moment in which we update the forecasts for GDP until the day it is realized, i.e. the release date.

In our set-up, the size and power of that test would be too much dependent on the number of times we update the model. We can update it every time we have a new data release, or update it every two weeks, for example. However, what is relevant for us is not whether the model produces rational multi-horizon forecasts, which is likely because they are based on a unique model with parameters obtained via maximum likelihood. Instead, we ask what are the forecasting updates that are most likely to yield significant improvements in forecasting accuracy. The results are available in Table A.2.

Table A.2: Statistical significance of each update based on fixed-smoothing (FS) asymptotics

Evaluation period: 2007.Q1 - 2015.Q1, T=25

	FS-Efficiency		**FS-DM**	**FS-Encompassing (U)pdate vs (O)ld**	
Real-Time Updates	bias	corr	RMSE	U enc O	O enc U
ARIMA	-0.27	0.50	-	-	-
DFM -90 (d)ays	-0.22	0.41	0.68	0.60	0.39
DFM -75 d	-0.19	0.47	0.55	-0.60	1.59
DFM -60 d	-0.12	0.55	0.52	0.26	0.51
DFM -45 d	-0.14	0.54	0.54	1.48	-0.54
DFM -30 d	-0.08	0.58	0.50	-0.20	1.07
DFM -15 d	-0.13	0.46	0.41	-0.65	1.59
DFM 0 d (end of quarter)	-0.06	0.45	0.38	-0.13	0.82
DFM +15 d	-0.09	-0.11	0.27	-0.02	1.01
DFM +30 d	-0.07	-0.08	0.26	-0.39	1.23
DFM +42 d	-0.10	-0.06	0.26	0.27	0.66
DFM +44 d	-0.06	-0.18	0.23	-0.17	1.03
BLOOM +44 d	-0.01	-0.20	0.17	0.23	0.69

Note: The FS-Efficiency multicolumn of his table reports bias and autocorrelation for the forecast errors obtained at different horizons. The FS-DM and FS-Encompassing blocks should be considered simultaneously. They aim to determine for each forecasting update (U) whether there is any added value with respect to the old/last available forecast (O). The null hypothesis of the Diebold-Mariano (DM) test is rejected when the *difference in the squared errors of U and O* is significantly different from zero. For the two encompassing tests, the null hypothesis states that the updated forecast (U) encompasses all the relevant information from the old forecast (O) (*or vice versa*). When the null hypothesis can be rejected, this implies that *U can be improved by combining it with O*. The combination weight associated to O (*or U*) is therefore reported below the "U enc O" test. In order to assess the added value of the updated forecast, the DM null of equal forecast accuracy should be rejected and at the same time the null "U enc O" and "O enc U" should be, respectively, not rejected and rejected. Given the small size of our evaluation sample and the time-series correlation patterns, we determine significance at the 5%, 10% and 20% level using the fixed-smoothing (FS) asymptotics, as proposed by Coroneo and Iacone (2016).

Table A.3: Our DFM compared to competitive benchmarks

Evaluation period: 2011.Q3 - 2015.Q1, T=15

Now-Casting.com (N-C)

	FS-Efficiency		**FS-DM**	**FS-Encompassing DFM vs Benchmark**	
Nowcasts	bias	corr	Rel RMSE	DFM enc Bench	Bench enc DFM
DFM -45 d	0.00	-0.05	1.21		0.34
N-C -45 d	0.02	-0.22		0.66	
DFM 0 d	0.11	0.04	1.31		0.33
N-C 0 d	0.00	-0.13		0.43	
DFM +44 d	0.00	0.14	0.70		0.56
N-C +44 d	-0.08	0.58		0.28	

Bloomberg (BLO) and Markit rule (PMI)

	FS-Efficiency		**FS-DM**	**FS-Encompassing DFM vs Benchmark**	
Nowcasts	bias	corr	Rel RMSE	DFM enc Bench	Bench enc DFM
DFM 0 d	0.11	0.04	0.83		0.68
PMI 0 d	0.08	0.11		0.31	
DFM +44 d	0.00	0.14	1.13		0.37
BLO +44 d	-0.02	-0.09		0.60	

Note: The FS-Efficiency multicolumn of his table reports bias and autocorrelation for the forecast errors obtained at different horizons. The FS-DM and FS-Encompassing blocks should be considered simultaneously. They aim to determine whether forecasts based on the DFM and the corresponding benchmarks are significantly different. The null hypothesis of the Diebold-Mariano (DM) test is rejected when the *difference in the RMSE* is significantly different from zero. In this table, the relative RMSE, defined as the RMSE of the DFM divided by the RMSE of the benchmark, will indicate that the forecast performance of the DFM is better than that of the benchmark when the fraction is smaller than one. For the two encompassing tests, we reject the null hypothesis that the DFM encompasses all the relevant information from the benchmark (*or vice versa*) when *the DFM can be improved by combining it with the benchmark*. The combination weight associated to the benchmark (*or DFM*) is therefore reported below the "DFM enc Bench" test. In order to assess the added value of the DFM, the DM null of equal forecast accuracy should be rejected and at the same time the null "DFM enc Bench" and "Bench enc DFM" should be, respectively, not rejected and rejected. Given the small size of our evaluation sample and the time-series correlation patterns, we determine significance at the 5%, 10% and 20% level using the fixed-smoothing (FS) asymptotics, as proposed by Coroneo and Iacone (2016).